

STIC-Biotech/ChemLib

61167

From: Pak, Michael  
Sent: Wednesday, February 27, 2002 10:15 AM  
To: STIC-Biotech/ChemLib  
Subject: 08/816,011 sequence search

Sequence search - 2 month amendment  
App. #: 08/816,011  
Result format: Paper.  
Title: potassium channels ...

Please search:

Search commercial and interference database.

SEQ ID NO:36 and 63.

Thanks,

Mike Pak

Michael Pak  
Art Unit 1646  
Mailbox: CM1, Rm. 10C00  
Office: CM1, Rm. 10E13  
703-305-7038

Michael Pak  
USPTO  
Art Unit 1646  
CM1; Rm. 10E13  
703-305-7038

*Handwritten notes:*  
CIP of PCT/US 55/14364 10/25/95  
CIP of 08/332,312 10/31/94 Pat # 5558,028  
no support

RECEIVED  
FEB 27 2002  
STIC

if Contact:  
Sheppard

Searcher: \_\_\_\_\_  
Phone: tel: 308-4499  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: 3/1/02  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:

NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST(where applic.)

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 16:59:59 ; Search time 25.85 seconds  
(without alignments)  
2455.791 Million cell updates/sec

Title: US-08-816-011f-63

Perfect score: 2294

Sequence: 1 MVTNRSNTYAYVEQERPRD.....LTGGLWPVVEHFVDRKPSIL 434

Scoring table: BLOSUM62

Gapop 10.0

Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_17:\*

1: sp-archaea:\*

2: sp-bacteria:\*

3: sp-fungi:\*

4: sp-human:\*

5: sp-invertebrate:\*

6: sp-mammal:\*

7: sp-mhc:\*

8: sp-organelle:\*

9: sp-phase:\*

10: sp-plant:\*

11: sp-rodent:\*

12: sp-virus:\*

13: sp-vertebrate:\*

14: sp-unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2265	98.0	434	5 Q19495	Q19495 caenorhabdi
2	802	35.0	418	5 Q9U318	Q9U318 caenorhabdi
3	716.5	31.2	461	5 Q23599	Q23599 caenorhabdi
4	706.5	30.8	450	5 Q33871	Q33871 caenorhabdi
5	645	28.1	458	5 Q9VMB6	Q9VMB6 drosophila
6	574.5	25.0	800	5 Q21145	Q21145 caenorhabdi
7	367.5	16.0	475	11 Q9JP65	Q9JP65 mus musculu
8	343.5	15.0	456	6 Q9UG85	Q9UG85 homo sapien
9	336.5	14.7	456	6 Q9UG85	Q9UG85 homo sapien
10	328.5	14.3	456	11 Q9JIT8	Q9JIT8 mus musculu
11	327.5	14.0	475	4 Q9B2D2	Q9B2D2 homo sapien
12	320.5	14.0	428	10 Q9M5X9	Q9M5X9 arabidopsis
13	313.5	13.7	428	10 Q64603	Q64603 arabidopsis
14	302.5	13.2	458	11 Q9JHF0	Q9JHF0 mus musculu
15	302.5	13.2	458	11 Q9JHF0	Q9JHF0 mus musculu
16	301.5	13.1	460	11 Q9JIM1	Q9JIM1 mus musculu
17	301.5	13.1	460	11 Q9JIM1	Q9JIM1 mus musculu
18	291.5	12.7	476	5 Q9NH21	Q9NH21 drosophila
19	261.5	11.4	397	4 Q9NUS9	Q9NUS9 homo sapien

20 249.5 10.9 408 10 Q9FWV1  
21 221.5 9.7 586 5 Q9VU20  
22 217.5 9.5 404 5 Q9VPP0  
23 197 8.6 497 5 Q9GTP5  
24 196.5 8.6 513 5 Q93690  
25 195.5 8.5 415 6 Q9BG84  
26 195 8.5 491 5 Q9G343  
27 187 8.2 491 5 Q9G269  
28 185.5 8.1 285 4 Q9BW12  
29 181 7.9 501 5 Q9N9R1  
30 180 7.8 143 5 Q20396  
31 170 7.4 499 5 Q9NBV4  
32 151 6.6 463 5 Q9U763  
33 147 6.4 417 10 Q9SR64  
34 144 6.3 418 10 Q9M0Y3  
35 142 6.2 462 5 Q9TVQ1  
36 138.5 6.0 418 10 Q9M0Y2  
37 136 5.9 463 5 Q9Y010  
38 135 5.9 422 5 Q9NIH8  
39 135 5.9 487 2 Q9CKV7  
40 133 5.8 463 5 Q9Y0H9  
41 132.5 5.8 502 5 Q9GTP4  
42 132 5.8 530 8 Q9B8V3  
43 131 5.7 347 8 Q9G510  
44 131 5.7 422 5 Q9NIH7  
45 128.5 5.6 382 10 Q9SY87

## ALIGNMENTS

RESULT 1  
Q19495 PRELIMINARY; PRT; 434 AA.  
ID Q19495  
AC Q19495;  
DT 01-NOV-1996 (TREMREL. 01, Created)  
DT 01-NOV-1996 (TREMREL. 01, Last sequence update)  
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)  
DE SIMILARITY TO HNP36 PROTEIN.  
GN F16H11.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson L.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
RA Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX Wu X.;  
RN [3]  
RP Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U55376; AAA98003.1; -;  
InterPro; IPR002259; DER\_Nucleoside\_tran.



Query Match		31.2%; Score 716.5; DB 5; Length 461;
Best Local Similarity		35.8%; Pred. No. 1.2e-39;
Matches 163; Conservative 86; Mismatches 167; Indels 39; Gaps 9;		
QY	7 SNTYAYEQEA-FPRDKYNIYVWLVILVGVGVLPLPWNMFITIAPEYVYVWF-KPD	59
DB	11 NKTKKVEEESPEDEKGNLVFYIILHGIGTLMPWNMLTISYDYFESYKMLANSTIDMD	70
QY	60 -----GVETWYKKEFGSLTIGSQLPNASINVFNLFLIAGPLIYRVFAPVCNIVNLTI	114
DB	71 TGVVTCYPTVYSSNFQSIASQVNLNLLNFIIVKGGGLASRI--TVGLSIVAVCV	128
QY	115 ILILVILEPTEDSMGFWFVTGLMATSINFSNGLYENSIVYGVGGDFPHYTGALLIGNN	174
DB	129 ITTMMFIYVETSWLGTGFTLTITIIIVLNGANGVYQNSIFGLASELPFKYTNNAVIIGN	188
QY	175 ICGLLITVVKI----GVTYFLNDEPKLVAIVYFGISLVILLVCAIALFFITKQDFYHYHQ	231
DB	189 LCGTFVTLISMSTKAVTRNIDR---SFAYFSIALITLVFCFISPHILKKQRFYQYST	244
QY	232 KGMEIREK---AETDRSPSILWTTFTNCYQGLFNWVFCFVAVTLTIFP-VMTVTVTRGDS	287
DB	245 RAERQRKNKDEAVDSGKVANIATFKEAPPQQLINLVFVFTLSIFPGVMVTKDEKKG	304
QY	288 G-----FLNKIMSENDEI---YTLTSLFVNLFAATIGSIVASKIHWPTPRYLK	333
DB	305 GYVDFPLPQQFINSLEIHNREFSEKYFMDVTTFLQFNVFAPFISIVAGRKQWPAPNKLW	364
QY	334 FAILRALRIPFFFCNRYVOTRAYVPVFESTDIFVIGGIMSFSHGYSALAMGYTPNV	393
DB	365 IPVYLLRLLYIPFIFCNYPETRSLSVPFFESTWLFVIIAASMSFGSGYTSGLAMYTSTK	424
QY	394 VPSHYSRFAAQLSVCTLVGLTGLGWPVYVIEHFV 428	
DB	425 VDPKQAQVAGMAGFFLSIGVSGLIFTWVVKWV 459	
RESULT	4	
Q93871	PRELIMINARY; PRT; 450 AA.	
AC	Q93871; (TremBLrel. 02, Created)	
DT	01-FEB-1997 (TremBLrel. 02, Last sequence update)	
DT	01-FEB-1997 (TremBLrel. 02, Last sequence update)	
DE	01-JUN-2001 (TremBLrel. 17, Last annotation update)	
DE	K09A9.3 PROTEIN.	
GN	K09A9.3	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;	
OC	Rhabditidae; Peloderinae; Caenorhabditis.	
OX	NCBI_TaxID=6239;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Swinnburne J.;	
RL	Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.	
RP	[2]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=94150718; PubMed=7906398;	
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,	
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,	
RA	Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,	
RA	Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,	
RA	Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,	
RA	Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,	
RA	Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,	
RA	Snaldon N., Smith A., Sonnhammer E., Staden K., Sulston J.,	
RA	Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,	
RA	Watson A., Weinstein L., Wilkinson-Sproat J., Wohlman P.;	
RT	*2.2 Mb of contiguous nucleotide sequence from chromosome III of C.	
RT	elegans *;	
RL	Nature 368.32-38(1994).	
DR	EMBL; Z79601; CAB01882.1; -	
DR	InterPro; IPR002259; DER_Nucleoside_tran.	
DR	Pfam; PF01733; Nucleoside_tran; 1.	
DR	ProDom; PD005103; DER_Nucleoside_tran; 1.	
SQ	SEQUENCE 450 AA; 50495 MW; CF3973D5EB517909 CRC64;	
Query Match		30.8%; Score 706.5; DB 5; Length 450;
Best Local Similarity		36.0%; Pred. No. 5.5e-39;
Matches 157; Conservative 88; Mismatches 164; Indels 27; Gaps 8;		
QY	12 VEQEA-FPRDKYNIYVWLVILVGVGVLPLPWNMFITIAPEYVYVWF-----KPD	59
DB	21 VEETPEPEKGNLVFYIILHGIGTLMPWNMLTISYDYFESYKMLANSTIDMDTGKVT	80
QY	60 GVETWYKKEFGSLTIGSQLPNASINVFNLFLIAGPLIYRVFAPVCNIVNLTIILV	119
DB	81 GDPVYSSNFQSIASQVNLNLLNFIIVKGGGLAGRI--TVGLSIVAVCVITMI	138
QY	120 IVLEPTEDSMGFWFVTGLMATSINFSNGLYENSIVYGVGGDFPHYTGALLIGNNCGLL	179
DB	139 FIYVETSTWLTGFTLTITIIIVLNGANGVYQNSIFGLASELPFKYTNNAVIIGNLCGT	198
QY	180 ITVVKI---GVTYFLNDEPKLVAIVYFGISLVILLVCAIALFFITKQDFYHYHKGMEI	236
DB	199 VTLSSTKAVTRNIDR---SFAYFSIALITLVFCFISPHILKKQRFYQYSTRAERQ	254
QY	237 R---EKAETDRSPSILWTTFTNCYQGLFNWVFCFVAVTLTIFP-VMTVTTRGDSGLNK	292
DB	255 RAKNEEAADNEGKMANVIATFKEAPPQQLINLVFVFTLSIFPGVMVTKDEKKGTYDF	314
QY	293 INSENDEITLTSFLVNLFAATIGSIVASKIHWPTPRYLKFAILRALRIPFFFCNRY	352
DB	315 PLPQN--YFMDVTTFLQFNVFAPFISIVAGRKQWPAPNKLWIPVYLLYIPFIFCNYL	372
QY	353 VQTRAYVPVFESTDIFVIGGIMSFSHGYSALAMGYTPNVPSHYSRFAAQLSVCTLMV	412
DB	373 PETRLPVFFESTWLFVIIAASMSFGSGYTSGLAMYTSTKTPDPSKAQVAGMAGFLIS	432
QY	413 GLTGLGWPVYVIEHFV 428	
DB	433 GIVSGLIFTWVVKWV 448	
RESULT	5	
Q9VMB6	PRELIMINARY; PRT; 458 AA.	
AC	Q9VMB6; (TremBLrel. 13, Created)	
DT	01-MAY-2000 (TremBLrel. 13, Last sequence update)	
DT	01-JUN-2001 (TremBLrel. 17, Last annotation update)	
DE	CG11045 PROTEIN.	
GN	BEST:LD04971 OR CG11045.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=BERKELEY;	
RX	MEDLINE=20196006; PubMed=10731132;	
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Ananidis P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	
RA	Abri J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,	
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,	
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
RA	de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	



RESULT	9	
Q9BG85		
ID	PRELIMINARY;	PRT; 456 AA.
AC	Q9BG85	
DT	01-JUN-2001 (TrEMBLrel. 17, Created)	
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)	
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)	
DE	NMNPR-INSENSITIVE NUCLEOSIDE TRANSPORTER EI.	
GN	ENT2.	
OS	Oryctolagus cuniculus (Rabbit).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;	
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctola	
OX	NCBI_TaxID=9986;	
RP	[1]	
RN	SEQUENCE FROM N.A.	
RA	Wu S.K., Ann D.K., Kim K., Lee V.H.L.;	
RT	"Molecular cloning and functional characterization	
RT	nucleoside transporters (rENT2 and its constitutive	
RT	rENT2A) in rabbit";	

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF323951; AAK11605.1; -  
SQ SEQUENCE 456 AA; 49858 MW; C6D55FC9D9D2A392 CRC64;

Query Match 14.7%; Score 336.5; DB 6; Length 456;  
Best Local Similarity 24.9%; Pred. No. 1e-14;  
Matches 115; Conservative 68; Mismatches 195; Indels 83; Gaps 13;  
Qy 18 PRDXYNIYVWLVILVGVGVLPPWNNFTIAPYVYVWYVWFKPDGVETWYSKEFMGS----- 72  
Dy 7 PRDSYHLVIGISFFILGLTLLPWNFFTAIP-YFQGRLAGANGAGALSTNHTGPTDAFN 65  
Qy 73 ----LTGSQLPNASINVENFLIAGPLIYRVFAPVCFNIVNLTIIILVIVLEPTDS 128  
Dy 66 FNNWVTLSQLPPLLFTLLNSFLYQCIPEAVRIIG-----SLAMLLLFALTAALVKVDVS 121  
Qy 129 MSWFFWVTLGMATSNFNSGLYNSVYGVGDFPHYIIGALLIGNNICGL-----LITVV 183  
Dy 122 PGPFESITMASAWFINSFCAVLQSLGQLGTMPSYNTLFLSGGLAGIPAALAMULTSM 181  
Qy 184 KIGVYFLNDEPKLVAIYVF---GISLVLLVCAIAFFITKQDFYHYHOK-----GM 234  
Dy 182 ASGV-----DAQTSALGYFLTPCGVILSVYLSLPHL---EFARYILAKPSQAPTQ 232  
Qy 235 EIREKAE-----TDRSPSILWTTFTNCY 258  
Dy 233 ELETKAELLQCKEKNVPSSPKAALTSVDVPEKEPELEPAEPDGPKPSV-FVVFRIKI 291  
Qy 259 GOLFNWFCFAVTLTIFFVMVTVTRGDSGLNKMSENDEIYTLTSLFVFNFAAIGS 318  
Dy 292 LTALCLVLVFTVLSVFPATAMVTSSTS-----PGKWSFEFFNPICCFLLFNVDMLGR 345  
Qy 319 IVASKIHWPP---TPRYLKFAIILRALFIPFFFCNRYVQTRAYVPFFESTDFIVGIGAMS 376  
Dy 346 SLTSYFLWVENSRLPLLRLFLVPLFVLMCHVPOHAR-LPIIFPQAAFIIFMLLFA 404  
Qy 377 FSHGYLSALAMGYTPNVVPSHYSFRAAQLSVCTLMVGLLTG 417  
Dy 405 VSNGLVSLTMCCLAPROVLAHEREVAGALMTFFFLALGLSCG 445

RESULT 10  
Q9JIT8  
ID Q9JIT8 PRELIMINARY; PRT; 456 AA.  
AC Q9JIT8  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE EQUILIBRATIVE NITROBENZYLTHIOINOSINE-INSENSITIVE NUCLEOSIDE  
DE TRANSPORTER ENT2.  
GN SLC29A2 OR ENT2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD1; TISSUE=WHOLE BRAIN;  
RA Kiss A., Farah K., Kim J., Garriock R., Drysdale T., Hammond J.R.;  
RT "Molecular cloning and functional characterization of inhibitor-  
sensitive and inhibitor-resistant equilibrative nucleoside transport  
proteins from mouse brain."  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF183397; AAF78477.1; -  
DR MGD; MGI:11345278; SLC29a2.  
DR InterPro; IPR002259; DER\_Nucleoside\_tran.  
DR Pfam; PF01733; Nucleoside\_tran; 1.  
DR PRINTS; PR01130; DERENTRNSPRT.  
DR ProDom; PD005103; DER\_Nucleoside\_tran; 1.  
SQ SEQUENCE 456 AA; 50255 MW; A9FE7CA037A4468F CRC64;

Query Match 14.3%; Score 328.5; DB 11; Length 456;  
Best Local Similarity 24.2%; Pred. No. 3.4e-14;  
Matches 113; Conservative 68; Mismatches 191; Indels 95; Gaps 14;  
Qy 18 PRDXYNIYVWLVILVGVGVLPPWNNFTIAPYVYVWYVWFKPDGVETWYSKEFMGS----- 72  
Dy 7 PRDSYHLVIGISFFILGLTLLPWNFFTAIP-----YFQGRLAGTNSAETMGNTHTSP 60  
Qy 73 -----LTGSQLPNASINVENFLIAGPLIYRVFAPVCFNIVNLTIIILVIVLE 123  
Dy 61 TDTNFNNWVTLSQLPPLLFTLLNSFLYQCIPESVRIIG-----SLAILLLFALTAALV 116  
Qy 124 PTEDSMWFFWVTLGMATSNFNSGLYNSVYGVGDFPHYIIGALLIGNNICGLITVV 183  
Dy 117 KVDLSPLGLFFSVTWASVWFINSFCAVLQSLGQLGTMPSYSTLFLSGGLAGIFAALA 176  
Qy 184 KI-----GVTYFLNDEPKLVAIYVF---GISLVLLVCAIAL-----FPI----- 220  
Dy 177 MLMSLASGV-----DAQTSALGYFITPCVIGILSVYLSLPHLKFARYLYTEKLSQAP 230  
Qy 221 ----TKODFYVHHKGMETR-----EKAETDRPSILWTTFTNCYG 259  
Dy 231 TQETLEKAEQLQADEKNGVPLSPQASPTLDLPEKEPEPEEPKPKPSVVFVFRKIWL 290  
Qy 260 QLFNWFCEFAVTLTIFFVM---MTVTTRGDSG---FLNKMSENDEIYTLTSLFVFNLF 313  
Dy 291 TALCLVLVFTVLSVFPATAMVTSSTSNGKGLFENPI-----CCLEFNVM 339  
Qy 314 AAGSIVASKIHWPP---TPRYLKFAIILRALFIPFFFCNRYVQTRAYVPFFESTDFIVI 370  
Dy 340 DWLGRSTYSFLWDEDSQQLPLLVLCLRLFLVPLFVLMCHVPOHAR-LPIIFRQDAYFIT 398  
Qy 371 GGIAWFSHGVSALAMGYTPNVVPSHYSFRAAQLSVCTLMVGLLTG 417  
Dy 399 FMLLFVSNGLVSLTMCCLAPROVLAHEREVAGALMTFFFLALGLSCG 445

RESULT 11  
Q9BZD2  
ID Q9BZD2 PRELIMINARY; PRT; 475 AA.  
AC Q9BZD2  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE EQUILIBRATIVE NUCLEOSIDE TRANSPORTER 3.  
GN ENT3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RA Hyde R.J., Cass C.E., Young J.D., Baldwin S.A.;  
RT "The ENT family of eukaryote nucleoside and nucleobase transporters:  
recent advances in the investigation of structure/function  
relationships and the identification of novel isoforms."  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF326987; AAK0958.1; -  
SQ SEQUENCE 475 AA; 51874 MW; 59089BB8A5C75FA2C CRC64;

Query Match 14.3%; Score 327.5; DB 4; Length 475;  
Best Local Similarity 24.4%; Pred. No. 4.1e-14;  
Matches 116; Conservative 83; Mismatches 202; Indels 75; Gaps 15;

Qy 7 SNTYAVEQEAPE-----PRDKYNIYVWLVILVGVGVLPPWNNFTIAPYV 52  
Dy 21 SSSLRADQEALEKLLDRPPGLQRPDRFCGTVIIFFSLGSLLPWNFFIT-AKEYWM 79  
Qy 53 NYWFK-----PDGVETWYSKEFMGSITIGSLPNASINVENFLIAGPLIYR 100  
Dy 80 ---FKLRNSSPATGEDPEGSD---ILNYFESYLAVASTVPSMLCLVANFLNVRVAVHIR 134



Db	183	GS	GVLS	LLRILTKAVY	PQDPDGL	RSKANLYFA	VGIVVMV	ICAVFYNA	IKLVP	IKFHEE	242	
Qy	232	KG	ME--	IREKAE----	TDRP	SESI	LLTTT	NCYGO	LNFNM	FCFAVTL	TIFFPMM	284
Db	243	RK	BEELIRE	KSEEGSL	TGLAWRT	TLWDI	VTVK	HSGF	IGVLL	YMTLS	IFPGYIT	298
Qy	285	GD	SGLN	KIMSE	NDIYT--	LLTSF-----	LVN	PLFA	AGTSI	VASKI	HNWTP	337
Db	299	----	----	----	----	----	----	----	----	----	----	345
Qy	338	LR	ALFI	PPFFC	NRVQ	TRAYP	VFVE	STDI	FV	GGIAMS	PHSGYLS	397
Db	346	AR	LLFY	PLF	WG	CLHG-----	PMEL	RT	ET	IPVTT	ILTCL	399
Qy	398	YS	RFAA	OLS	VCTLM	VGL	LLT	GG	L--	WP	VI	424
Db	400	HS	ETAG	IVT	VTFL	VVG	LAGS	VT	AW	FW	VI	428

RESULT	13
ID	064603
ID	064603
AC	064603;
DT	01-AUG-1998 (TREMBlrel. 07, Created)
DT	01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)
PRELIMINARY;	PRT; 428 AA.

OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Vysotskaja V.S., Schwartz J.R., Toriumi M., Yu G., Kwan A., Oji O.,  
RA Liu S., Li J., Araujo R., Au M., Brendel V., Buehler E., Conway A.B.,  
RA Conway A.B., Dewar K., Feng J., Kim C., Kurtz D., Li Y., Palm C.J.,  
RA Shinn P., Sun H., Davis R.W., Ecker J.R., Federspiel N.A.,  
RA Theologis A;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Theologis A;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Theologis A;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Theologis A;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Theologis A;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC003671; AAC18807.1; -;  
DR InterPro; IPR002106; AA-TRNA\_ligase\_II.  
DR InterPro; IPR002259; DER\_Nucleoside\_tran.  
DR Pfam; PF01733; Nucleoside\_tran; 1.  
DR ProDom; PD005103; DER\_Nucleoside\_tran; 1.  
DR PROSITE; PS00339; AA-TRNA\_LIGASE\_II\_2; UNKNOWN\_1.  
SQ SEQUENCE 428 AA; 46764 MW; B762E36C07516DB9 CRC64;

```

Query Match      13.7%  Score 313.5;  DB 10;  Length 428;
Best Local Similarity 23.2%;  pred. No. 3.1e-13;
Matches 100;  Conservative 84;  Mismatches 186;  Indels 61;  Gaps 11;

Qy 18 PRDRINYVWLIVLGVGLLPPNMNFITIAPEYVYVNWFKPDGVETYSKREFMGLSTIGS 77
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 35 PSDSYHEAYIIYFLGVGFLPPNNAFTAVD--YFSVLYPSTAVDRIFAVIYM----- 85

```



Search completed: February 27, 2002, 17:03:21  
Job time: 202 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2002, 23:22:38 ; Search time 1347.44 Seconds  
(without alignments)  
16993.770 Million cell updates/sec

Title: US-08-816-011f-36  
Perfect score: 1388  
Sequence: 1 atggttaataatgaacgatc.....ttattaaaaaaaaaaaaa 1388

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues  
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenBank:

1: gb\_ba:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pi:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vi:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_om:\*

20: em\_or:\*

21: em\_ov:\*

22: em\_pat:\*

23: em\_ph:\*

24: em\_pi:\*

25: em\_ro:\*

26: em\_sts:\*

27: em\_sy:\*

28: em\_un:\*

29: em\_vi:\*

30: em\_htgo\_hum:\*

31: em\_htgo\_inv:\*

32: em\_htgo\_rod:\*

33: em\_htg\_hum:\*

34: em\_htg\_inv:\*

35: em\_htg\_rod:\*

36: em\_htg\_other:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	101.2	72.9	31049	3	U55376	U55376 Caenorhabdi
c 2	90.8	6.5	13429	2	AC019610	AC019610 Drosophil
c 3	90.8	6.5	64674	3	AC004247	AC004247 Drosophil
c 4	90.8	6.5	174376	3	AC092230	AC092230 Drosophil
c 5	90.8	6.5	259464	3	AE003614	AE003614 Drosophil
c 6	84.8	6.1	31853	3	CEFA4D12	282263 Caenorhabdi
c 7	71.2	5.1	40664	3	CEFA4784	282263 Caenorhabdi
c 8	70.8	5.1	25560	3	CEZK809	268303 Caenorhabdi
c 9	68	4.7	7218	6	I66494	I66494 Sequence 14
c 10	65.8	4.7	34017	3	CEK09A9	Z79601 Caenorhabdi
c 11	47.8	3.4	253305	3	PFMAL3P7	AL034559 Plasmodiu
c 12	44.6	3.2	3885	3	AF151733	AF151733 Dictyoste
c 13	44.6	3.2	3985	3	AF151111	AF151111 Dictyoste
c 14	44.6	3.2	110211	8	AC005170	AC005170 Arabidops
c 15	43.8	3.2	366	8	SCOR17	X59536 S.cerevisia
c 16	43.8	3.2	2468	8	YSCMTORIK	M11449 Yeast mitoc
c 17	43.8	3.2	21153	8	YSCMTG13	L36897 Saccharomyc
c 18	43.8	3.2	85779	8	SCE011856	AJ011856 Saccharom
c 19	43.8	3.2	112576	2	AC011250	AC011250 Drosophil
c 20	43.8	3.2	184621	3	AC007440	AC007440 Drosophil
c 21	43.8	3.2	262731	3	AE003823	AE003823 Drosophil
c 22	43.4	3.1	77835	2	PFMAL13P2_3	Continuation (4 of
c 23	43	3.1	1141	6	AX083744	AX083744 Sequence
c 24	43	3.1	33637	3	CEY10G11	AL161711 Caenorhab
c 25	42.8	3.1	175737	9	HS1033B10	AL031228 Human DNA
c 26	42.4	3.1	2175	3	AF139082	AF139082 Haematobi
c 27	42.4	3.1	321003	2	PFMAL4P3	AL035476 Plasmodiu
c 28	42.2	3.0	104992	2	AC005504	AC005504 Plasmodiu
c 29	42.2	3.0	169546	2	AC004157	AC004157 Plasmodiu
c 30	42	3.0	23269	8	SPAC1856	AL391783 S.pombe c
c 31	42	3.0	23269	8	SPAC1856	AL391783 S.pombe c
c 32	41.6	3.0	175438	2	AC084352	AC084352 Homo sapi
c 33	41.6	3.0	321003	2	PFMAL4P3	AL035476 Plasmodiu
c 34	41.2	3.0	64248	2	AC017951	AC017951 Drosophil
c 35	41.2	3.0	84055	3	AC004439	AC004439 Drosophil
c 36	41.2	3.0	170868	3	AC092221	AC092221 Drosophil
c 37	41.2	3.0	306352	3	AE003586	AE003586 Drosophil
c 38	41	3.0	315	3	ECA270225	AJ270225 Entodiniu
c 39	40.8	2.9	178586	2	AC073485	AC073485 Homo sapi
c 40	40.8	2.9	191122	9	AC012494	AC012494 Homo sapi
c 41	40.6	2.9	14923	3	AE001394	AE001394 Plasmodiu
c 42	40.6	2.9	88512	8	F21H2	AC007894 Arabidops
c 43	40.6	2.9	110619	8	AC018460	AC018460 Arabidops
c 44	40.6	2.9	199551	2	AC006281	AC006281 Plasmodiu
c 45	40.4	2.9	2206	8	SCYK072C	Z28297 S.cerevisia

## ALIGNMENTS

RESULT 1	U55376/c	31049 bp	DNA	INV	09-AUG-2001
LOCUS	U55376	Caenorhabditis elegans cosmid F16H11, complete sequence.			
DEFINITION	U55376				
ACCESSION	U55376.1	GI:1280130			
VERSION	HTG.				
KEYWORDS	HTG.				
SOURCE	Caenorhabditis elegans.				
ORGANISM	Caenorhabditis elegans				
REFERENCE	1 (bases 1 to 31049)				
AUTHORS	The C. elegans Sequencing Consortium.				
TITLE	Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium				
JOURNAL	Science. 282 (5396), 2012-2018 (1998)				
MEDLINE	99069613				
REFERENCE	2 (bases 1 to 31049)				
AUTHORS	Wu.X.				
TITLE	The sequence of C. elegans cosmid F16H11				

JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 31049)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 4 (bases 1 to 31049)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-APR-1996) Robert Waterston  
 REFERENCE 5 (bases 1 to 31049)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-AUG-2001) Department of Genetics, Washington  
 REFERENCE University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
 JOURNAL Louis, MO 63110, USA  
 COMMENT Submitted by:  
 Genome Sequencing Center  
 Department of Genetics, Washington University  
 St. Louis, MO 63110, USA, and  
 Sanger Centre, Hinxton Hall  
 Cambridge CB10 1RO, England  
 email: rw@ematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.  
 It may be shorter because we only sequence overlapping sections  
 once, or longer because we provide a small overlap between  
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all  
 regions were double stranded, sequenced with an alternate chemistry  
 or covered by high quality data (i.e., phred quality >= 30); an  
 attempt was made to resolve all sequencing problems, such as  
 compressions and repeats; all regions were covered by sequence from  
 more than one m13 subclone.

#### NEIGHBORING COSMID INFORMATION

The 5' cosmid is C46C11, 200 bp overlap; 3' cosmid is K02G10, 2500  
 bp overlap. Actual start of this cosmid is at base position 1 of  
 CELF16H11; actual end is at 10112 of CELK02G10.

#### NOTES:

Coding sequences below are predicted from computer analysis, using  
 the program Genefinder(P. Green and L. Hillier, ms in preparation).

FEATURES  
 source  
 1. .31049  
 /organism="Caenorhabditis elegans"  
 /strain="Bristol N2"  
 /db\_xref="taxon:6239"  
 /chromosome="X"  
 /clone="F16H11"  
 /complement(1239..2839)  
 /gene="F16H11.3"  
 complement(join(1239..1368,1421..1617,1669..1878,  
 1932..2131,2180..2466,2516..2649,2693..2839))  
 /gene="F16H11.3"  
 /note="weak similarity to HNP36 protein"  
 /codon\_start=1  
 /product="Hypothetical protein F16H11.3"  
 /protein\_id="AAA98003.1"  
 /db\_xref="GI:1280131"  
 /translation="MVIIKSNPTAYVEQEAFFPRDKYNIYVILVGFGLPWNMF1  
 TIAPEYVNVFPCDGVETWYKEFMGSLT IASQLPNASINVENFLIAGPLIYRVF  
 APVCENIVNTITILVILEPTEDSMFFFWTLGMATSIENSGLYENSVYGVAD  
 PPHYTGALLIGNICGLLITVYKIGTYVFLNDPEKLVAIVYGISILVILLVCAIALF  
 FITQDFYHYHKGMEIREKAETDRPSILMTFTNCTGQLFNWVFCFVLTLPF  
 VMVTTRGDSGLNKIMESENDEIYLLTSFLVFNFAIGSIVASKIHWPTPYRIKE  
 AIIIRALFTPFPPFCINRYQTRAYPVFFESTDIFVIGGIAMSESHGYLSALAMGYTPN  
 VVPSHYSRFAQLSVCTLMVGLLTGGLWAVVIEHFVDKPSIL"  
 complement(7137..7683)  
 /gene="F16H11.4"  
 /note="ceh-1"

CDS  
 complement(join(7137..7253,7307..7456,7505..7573,  
 7621..7683))  
 /gene="F16H11.4"  
 /note="C. elegans unknown homeobox gene ceh-1 (GB:X52810)"  
 /codon\_start=1  
 /product="Hypothetical protein F16H11.4"  
 /protein\_id="AAA98004.1"  
 /db\_xref="GI:1280132"  
 /translation="MRRARTAFYEQLVLENKFKTSRYLSVVERLNLAIQLOLSETQ  
 VKTFONRRTKKKHNPQODANTPOTPPSSDETQIQIPILPANPITSSSSLLLPITSP  
 ANSAVQGTSPITLNLNQLMPQNVGFN"  
 complement(10470..14006)  
 /gene="F16H11.5"  
 /note="nhr-45"  
 complement(join(10470..10481,10529..10624,10701..10861,  
 10907..11025,11086..11396,11509..11613,12150..12431,  
 13147..13430,13566..13631,13796..14006))  
 /gene="F16H11.5"  
 /note="contains a zinc finger of the C4-type; coded for by  
 the following C. elegans cDNAs: yk347g3.3, yk347g3.5"  
 /codon\_start=1  
 /product="Hypothetical protein F16H11.5"  
 /protein\_id="AAA98005.1"  
 /db\_xref="GI:1280133"  
 /translation="MTHTFSINTLFFSETTSKISMSCLVCETAHQHFGIRCCRAC  
 AAPFRRTLMNRYKCFDRKCEVSNKRYSCRCRYEKCVRYGMRKDNLTSPVSVVQ  
 NKSETEKDSNETSDIGSHSPSSLESYTCRPPSFEQQQONQPDPLLMNGNVTLLS  
 EVASDAYQNSDMHOLFSEFHRTYIEEVGRAVMPPTFPQOPPIHIPOPDFOYVTLTLLS  
 TDQNSASPSSTOYNOVQYENNESGNQNMFAQAQATODLHVNQVNLTPITD  
 AMFGPPDAFCQLPDILPLCQALLAYREHNKQWQDQKNIENVPDLMENFNRNHYI  
 EIEHIAFCMSIRVFAQLPKDQKIIFKHFWTRFYELDRFCATCQRLGYNLTERGTL  
 LTIHINFGISVVKLEISMDATQWVNFELKSMDFRLIFINPKFKLQPTTELYMVM  
 MMSIMSVSNLPGITDRTDISKVELLAEDLHTYAEQYDNNPNYAGRITRLSSI  
 ASAVDIEKTKRSDSQVSKTFNFKNDFEFDLTDFFCLN"  
 16440..16613  
 /gene="F16H11.2"  
 16440..16613  
 /gene="F16H11.2"  
 /codon\_start=1  
 /product="Hypothetical protein F16H11.2"  
 /protein\_id="AAA98006.1"  
 /db\_xref="GI:1280134"  
 /translation="MTQRDRTEERERERERERERERERERERERERERERE  
 REREREKESNRRK"  
 21734..26707  
 /gene="F16H11.1"  
 join(21734..21787,21843..21909,21958..22079,24004..24109,  
 24156..24295,24739..24809,24855..24966,25036..25127,  
 25173..25257,25305..25397,25441..25536,25584..25751,  
 26297..26394,26451..26593,26649..26707)  
 /gene="F16H11.1"  
 /note="similar to melibiose carrier protein  
 (thiomethylgalactoside permease II); coded for by the  
 following C. elegans cDNAs: cm01e2, cm21e6"  
 /codon\_start=1  
 /product="Hypothetical protein F16H11.1"  
 /protein\_id="AAA98007.1"  
 /db\_xref="GI:1280135"  
 /translation="MTDETEGDVVPVRSDDIVIPRRSIRPDHDSFTKKIAFGIGH  
 FYNDCASWFTYFIYMEVKLFQFSRAGMLMIGVTDASTPLVGLFSDSNILPA  
 CFDKIGRMSWHLIGTVLSLSPFMIFNCKFCICKSTISWLVKLVFVFMVTFQSWA  
 SVOISHLALIPELSSVPASRTMNSURAFYVIANLSVYPALAWLLESSTGHSTIGPW  
 DSHFSLAGLVVVLGVAFYAFRTPTNYRRESLNSFSSDASSELVRMTISGW  
 GHQVQIALLNLSRLYINISQVYFPFTYMTQYVYKRYALPLPVAQVILSSFSVSW  
 NSLPVYSKSKLTYSLFGLASMLSCAVNMLDLPWKIYALVAGVIAQAILTISLS  
 ITADLINKTESCAFYVYGMSPFDKLSNGTAVOLJELWTPAYDALKPHVEVSAFYRRV  
 MWPVGTGLVLAFLVLSLAPFKIGERRARPEQAINEDQDDIYPIVE"  
 BASE COUNT 10124 a 5100 c 5591 g 10234 t  
 ORIGIN  
 Query Match 72.9%; Score 1012; DB 3; Length 31049;  
 Best Local Similarity 82.0%; Pred. No. 3,3e-251;

Matches 1371; Conservative 0; Mismatches 5; Indels 296; Gaps 6;									
Qy	1	atgtataatcaaccgacgaacacacatcctgcttgaagcaggaagcatttccaaagaac	60						
Db	2839	ATGGTAATAATCAACCGATCAACACCTATGCCGTTGAGCAGGAGCAATTTCCAGAGAC	2780						
Qy	61	aagtacaatattgtctactggtcgtgctcattcttctgtgattcggagttctctcgcattg	120						
Db	2779	AAGTACAATAATTGCTACTGCTGCTCATTTCTGTTGGATTCCGAGTTCTTCTGCCATGG	2720						
Qy	121	aatatgttcattactatgcgcctg-----	145						
Db	2719	AATATGTTTCATTACTATATGCCCTCGAGGTTGCAACTAGATAAATTTTCACTAATAGAGTT	2660						
Qy	146	-----agtatatgtgaattatctgttcaaacccgattgctgagagcatgattc	197						
Db	2659	CCTGTTTCAGTATATGTGAATATTTGTTCAACCCGATGGCGTGGAGACATGGTATTC	2600						
Qy	198	gaaagaattcatgggatctttgagattggctcacaacttccaaacgcaagcattaatgt	257						
Db	2599	GAAGAATTCAATGGGATCTTTGAGATTCCTCACAACCTTCCAAACGCAAGCATTAATGT	2540						
Qy	258	tttcaacctgttccctcaattatgc-----	281						
Db	2539	TTTCAACCTGTTCCCTCATATTTCGCTAAGTTACAGTTCTTTAAACTCTCAAAAATCAAT	2480						
Qy	282	-----tqgtccctgatctaccgctcttctgctccggtttgtcttcaacatcg	328						
Db	2479	CAATATTTTTTAGTGGTCCCTGATCTACCGGCTTTGCTCCGGTTTGGCTTCAACATCG	2420						
Qy	329	tcaacctgacaatcatctcatcctcgtcattgttctggagcccactgaagattccattgt	388						
Db	2419	TCAACCTGACAATCATCTCATCTCTCGTCATTGTTCTGAGGCCACTGAAGATTCCATGT	2360						
Qy	389	cctggttttctgggaactcttggattggagcttcaactcaatttttagaatgggctat	448						
Db	2359	CCTGGTTTTCTGGTAACTCTTGAATGGCGACTTCAATCAATTTTAGCAATGGGCTAT	2300						
Qy	449	atgaaaactcgttttatgagttggtgagctatttccgcacacactacattggcgctctct	508						
Db	2299	ATGAAAACCTCGTATTGAGGATTTTGGCGATTTTCCGCACACCTACATTGGCGCTCTCT	2240						
Qy	509	tgattggaaaacacatttcgaggtatgctgataaacggttgtgaaatcggagtgacctatt	568						
Db	2239	TGATTGGAACAACAATTTGCGGATTCGTGATACGGTTGTGAAAATCGAGTACCTATT	2180						
Qy	569	-----tctgaatgatg	580						
Db	2179	GTAAGTAATTTTCAATTTATTTGCTTTTAAAACTATAACATTTTATGTTCTGAATGATG	2120						
Qy	581	agcctaaacttctgcaatcgtctatttccgcatatcgttggatccttctggtggtg	640						
Db	2119	AGCCTAAACTTTTGCATCTGCTATTTTCGCATATCTGTTGGTATCTCTTCTGTTGCTGTG	2060						
Qy	641	caattgcactttctttatcaaaagcaagattttaccacatatcaccatcaaaaaggaa	700						
Db	2059	CAATTGCACTTTCTTTATCAAAAGCAAGATTTCTACCACTATCACCATCAAAAAGGAA	2000						
Qy	701	tggaaattcggaaaaggcgaaccgacagacgctctccatcattcttggaccacat	760						
Db	1999	TGGAAATTCGGGAAAAGCGGAAACCGAGACCGCTCCCATCTCTTTTGGACCCACAT	1940						
Qy	761	tcacaaac-----	768						
Db	1939	TCACAAACGCTGAGTTTTTAAATCAATTTAGCTTCAAAATAAATATGTGACTAATTTCA	1880						
Qy	769	-tggtatgggcaactcttcaatgtttgtctgttgcgttactcttcaaatcttccc	827						
Db	1879	GTGTTATGGCAACTCTCAATGTTTGTGTTCTGCTTTCCGCTTACTCTCAACAATCTCCC	1820						
Qy	828	tgtatgatgacggttaccactcgtggagattccggttctcctaaacaaaatatgtctga	887						
Db	1819	TGTTATGATACCGTTACCACTCGTGGAGATTCCGGCTTCCTAAACAAAATATATGCTGA	1760						

RESULT 2

AC019610	13429 bp	DNA	HTG	03-JAN-2000
LOCUS	Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***			
DEFINITION	pieces.			
ACCESSION	AC019610			
VERSION	AC019610.1	GI:6665287		
KEYWORDS	HTG; HTGS_PHASE2.			
SOURCE	fruit fly.			
ORGANISM	Drosophila melanogaster			
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
AUTHORS	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;			
TITLE	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
JOURNAL	Adams M. and Venter, J.C.			
COMMENT	Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA			
	This sequence was identified as CDM:10210382 by the submitter. For more information on this record e-mail to fly@celera.com.			
	* NOTE: This is a 'working draft' sequence.			
	* This sequence will be replaced			
	* by the finished sequence as soon as it is available and			
	* the accession number will be preserved.			
FEATURES	Location/Qualifiers			
source	1. 13429			
	/organism="Drosophila melanogaster"			
	/db_xref="taxon:7227"			
BASE COUNT	3597 a 3053 c 2937 g 3842 t			
ORIGIN				

Query Match 6.5%; Score 90.8; DB 2; Length 13429;  
 Best Local Similarity 45.4%; Pred. No. 1.2e-12;  
 Matches 556; Conservative 0; Mismatches 632; Indels 36; Gaps 5;

QY	47	catttccaagagacaagtacaataattctacttactgctgcctcattctgttgattcgag	106
DB	7103	CAGCGCCAAAGATAAATCCTTATCGTGTCCTTATATTCCTGTGACCGCGTGGCA	7162
QY	107	tctctcgcatggaaatgtcttactacttcgcccc-----tgatttatgtgaatt	160
DB	7163	CGCTGATGCCATGAACATGTTTCATCAGGCCAAGTCTATTTCGAGGACTTCAGTTCG	7222
QY	161	atttggttcaaacggatggtcggtggagacatggtattcgaagaattcatgggactttga	220
DB	7223	GACCAAAACACACCGTGGCCAGCAAGTAGTTACCGCACCCCATTTTCATGCAAAATATGG	7282
QY	221	cgattggctcacaaactccaaacgcaagcattaatgtttcacaactgttctctcattatg	280
DB	7283	GCATTGCCTCGCAGATTCCTTAACCTGGTCTTCACTGGCTGAACATCTTTGTCACATTTG	7342
QY	281	ctggtccctctgatatcaccggtcttttgcctcgggttcttcaacaatcgtcaacctgacaa	340
DB	7343	CGCGGACCTTGACACCCGAATCGTCTACAGCATCATCTTCGAAATGCTCATCTGTGTGG	7402
QY	341	tcaattctacactcctgtcattgttctggagccccactgaagattccatgtcctcgttttct	400
DB	7403	TTACCATTATTTGGCCATGCTCG-----ACTCCTCGAGTGGCGGGCTCTCTCTCT	7456
QY	401	gggttaactcttggaaatggcgacttcaatcaatttagcaatgggcttatgaaactcgg	460
DB	7457	GGACCACCATGTGTGCATTGTGTGTTAAACGTGTGCAACGGCATCTTACCAGAACACCA	7516
QY	461	tttatggagtgtggtggatatttccgcacacctacattggcgctctcttggatggaaaca	520
DB	7517	TATACGAATACGTGCNCATCTGTGCCCATCAAGTACACAGGCGCGCTGTCCTGGGTCCA	7576
QY	521	acatttggcgattgctgataaacggttgtgaaaaatcggagtgacctattttctgaatgatg	580
DB	7577	ACATCAGCGGCTGTTCCACCACCGCCATGGCTTGATGCGGTGAGATCTTCTCGTCCA	7636
QY	581	agcctaaacttgtgaaatcgtctatttcggcatacttggtggatcctcttggtggtg	640
DB	7637	AG---AGGAGCTTCGGGCATCTACTACTCTGTGACGGGCATCCTAGTGTCTGTCTGCT	7693
QY	641	caatttgcactttctttatcacaaagcaagatttctaccactatcaccaatcaaaaagaa	700
DB	7694	TCGACACGTACTTTGGCGCTCCCGCTGAAACAAGTCTTCCGGCACTACGAGACCATTAGCC	7753
QY	701	tggaaatttcgaaaaaggcggaacacgacagacgtctccatccatcttttggaccacat	760
DB	7754	GGAGCAGCGAAAAGAAATCTGATTCTTAAGGCCACAGCTGAACGTGCTTACTGGCAGATCT	7813
QY	761	tcacaaactgttatgggcaactcttcaatgttggttctgtcttggcgttactctcaaa	820
DB	7814	TCAAGAAACGAGCGCCCACTATTCAACATCTTCCTCAGTTTTTTCGTCAGCTTGTGG	7873
QY	821	tcttccctgttatgtagaccgttaccactcgtggagattccggcttctcctaaacaaatta	880
DB	7874	TTTTCCCGGGATTCAGTCGAAGGTGCACCGCTCCGACCCAAAATTTTCGTA-----	7923
QY	881	tgtctgaaaaacgatgaaatctacactttgctcacaagtttctcgttctcgaattgtcg	940
DB	7924	----GTAGGCCCGGACTACTTTACACTGTGTCACTTCTCTTCGGGACCTTTAACGTC	7978
QY	941	ctgcgattggatccaagtttgcttccaaagattcaactggcgacaccccggttacctcaaat	1000
DB	7979	CAATGCTGGGACGTTTAAACACATCGTGGGTGCAATGGCGCGGTCCGAGGTTCTCTCTGGG	8038
QY	1001	ttgccaataacttgcgtgctcttttcaattccaattcttctcttcttgcacata-----tc	1054
DB	8039	TGCCAGTGTGTCGCTGTGGCGCTCATCCCTTTGTTGTCATGTGCAATACGATCGTCCGC	8098

Qy	1055	gtgtccagacggtgcttatactctgttttttgaagtctactgacatttttgaattggtg	1114
Db	8099	CGGACTCGGTGGCGTCATTGGGCGGTATTTCATCGAAGCAATTTGGGTGTACTCGGGGATCG	8158
Qy	1115	gaattgccaatgctttttcacatgatacctcagcgtctggcaatgggatacacatcccaa	1174
Db	8159	GCATCGCAATGGCGCTACAGTCCGGGTAFCTCAGCTCTCTGGGCATGATGTCAGCCCCGC	8218
Qy	1175	acgtgtgccatctcactactcaagaattggcgcgtcagctttccgtttgcactctttatgg	1234
Db	8219	AGACGGTCACACCAAAATACAGACCACTGCGGGAATGTACCGAGCTGCCATGCTATCA	8278
Qy	1235	ttggcctcttcacgggtggcctgt	1258
Db	8279	CGGGCATCTTCTCGGAGTGTGT	8302
RESULT	3		
LOCUS	AC004247/c	64674 bp	DNA
DEFINITION	Drosophila melanogaster (Pl DS07773 (D154))	INV	01-MAR-1998
ACCESSION	AC004247	AC003251	AC003249
VERSION	AC003761	AC003250	AC003248
KEYWORDS	HTG.	GI:2920811	
SOURCE	Drosophila melanogaster (Subclones in Sac from Pl clone DS07773 (D154))	DNA.	
ORGANISM	Drosophila melanogaster		
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
AUTHORS	1 (bases 1 to 64674) Celisner,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazef,R.G., Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B., Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacible,J.M., Park,S., Pfeiffer,B., Punch,D., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.		
TITLE	Sequencing of Drosophila chromosome 2L, region 26F1-26F4		

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 Sequencing of Drosophila chromosome 2L, region 26F1-26F4  
 Unpublished (1997)  
 2 (bases 1 to 64674)  
 Celnikier,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
 Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,  
 Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,  
 Lomoton,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M.,  
 Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., PUNCH,D., Snir,E.,  
 Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L.  
 and Kimmel,B.  
 Direct Submission  
 Submitted (28-FEB-1998) Berkeley Drosophila Genome Project, MS  
 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,  
 Berkeley, CA 94720, US  
 Sequence submitted by:  
 Lawrence Berkeley National Laboratory, MS 64-121  
 Berkeley, CA 94720  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our sequence  
 archive Web site  
 (<http://fruitfly.berkeley.edu/sequence-archive.html>) or send email  
 to [drosophila@genome.lbl.gov](mailto:drosophila@genome.lbl.gov).  
 Library location: 136.81.  
 This P1 was assembled from the following subclones: 1\_b4, 2\_b6,  
 2\_d8, 1\_h5 (AC003251), 1\_c11, 2\_e1, 1\_f2, 1\_g4 (AC003249), 1\_g6  
 (AC003250), 2\_c9 (AC003252), 1\_f12 (AC003248), 1\_b9, 2\_e9, 1\_a5,  
 2\_e3, 1\_e12 (AC003247), 2\_h1, 2\_b12, 1\_a12 (AC003761), 1\_g3, 1\_g12  
 2\_e7, 1\_c8.  
 Location/Qualifiers  
 1. 64674  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 FEATURES  
 source



Db	58871	CAATCTGGGCGAGTTTAAACACAATCTGTGGGTGCATAATGCCGGTGCCAGGTTCTCTGGG	58812
Qy	1001	ttagcataaatcttcggtgccttttccattcactctctcttcttgcaaca-----tc	1054
Db	58811	TGCCAGTGGTCTTCGCTGTGCGGTTCATTCCCTTGTGTCATGCAACTACGTTCCGC	58752
Qy	1055	gtatcgagacggctgtcttatctgttttttttgttgagtctactgacattttgtgatagggtg	1114
Db	58751	CGGAATCGGTGGCCTCATATGGCCGTAATCAFCGAGAAGCAATGCGGTGTACTGGGGCATCG	58692
Qy	1115	gaattgccatgctcttttccatcatgataacctcacgcctctgccaatgggatcacactccaa	1174
Db	58691	GCAATCGGAATGGCCTACAGTCCGGGTATCTCACGCTCTCTGGGATGATGTACC GCCCGC	58632
Qy	1175	acgtctgaccatctcactactaacagatttgcgcctcagctttccgtttgcaactcttatgg	1234
Db	58631	AGACGGTGCACACCAATACAGACCACTGCCGGAATCTACGCAGCTCCCATGCTGATCA	58572
Qy	1235	ttagcctcttcacccggtggcctgt	1258
Db	58571	CGGCGATCTTCTCCGAGTGTGT	58548
RESULT	4		
AC092230	LOCUS	174376 bp DNA INV	30-JUN-2001
DEFINITION	Drosophila melanogaster, chromosome 2L, region 26F-27X, BAC clone		
ACCESSION	BACR25K24, complete sequence.		
VERSION	AC092230		
KEYWORDS	HTG.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;			
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
REFERENCE	1 (bases 1 to 174376)		
AUTHORS	Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Ananides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Champagne,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houck,J.J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svirkas,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.		
JOURNAL	Sequencing of Drosophila chromosome 2L, region 26F-27X		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 174376)		
	Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Ananides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Champagne,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svirkas,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (30-JUN-2001) Berkeley Drosophila Genome Project, MS		
REFERENCE	64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,		
AUTHORS	Berkeley, CA 94720 US		
COMMENT	Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory, MS 64-121		

Berkeley, CA 94720  
This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our send email archive web site (<http://www.fruitfly.org/sequence/>) or send email to [bdgpe@fruitfly.berkeley.edu](mailto:bdgpe@fruitfly.berkeley.edu).

# FEATURES

source  
1. 174376  
/organism="Drosophila melanogaster"  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="2L"  
/map="26F-27X"  
/clone="BACR25K24 (D1294)"  
/clone\_lib="RPCI-98 (Roswell Park Cancer Institute  
Drosophila melanogaster BAC library, partial EcoRI in  
pBACe3.6)"  
BASE COUNT 48681 a 38138 c 38499 g 49058 t  
ORIGIN

Query Match 6.5%; Score 90.8; DB 3; Length 174376;  
Best Local Similarity 45.4%; Pred. No. 1.7e-12;  
Matches 556; Conservative 0; Mismatches 632; Indels 36; Gaps 5;  
QY 47 catttccaagagacaagtacaattatgtctacttggtcgctcattcttctgttgattcgag 106  
DB 39300 CAGCGCCAAAGATTAATCTTATCTGTTCTTTATTCCTGCTGACGCGGTGGCA 39359  
QY 107 ttcttgcctgaatgaattgttctacttactatcgcccc-----tgagtattatgtaatt 160  
DB 39360 CGCTGATGCCATGAACATGTTTCATCAGCGCAAGTCTTATTCGAGGACTTCAAGTTCG 39419  
QY 161 attggttcaaacggatg9cgtggagacatggtattggaagaattcaatggtatcttga 220  
DB 39420 GACCAACAAACACCGTGGCCAGCGAGTGAGTTACCGCACCCCATTTTCATGCAAAATATGG 39479  
QY 221 cgattgctcacaaactccaaacgcaagcattatgttttcaacctgttctcattatg 280  
DB 39480 GCTTTGGCTCGCAGATCTCTTAACCTGTGTCACACTGGCTGACACATCTTTGTCAACTTGG 39539  
QY 281 ctggttccctgatctaacgcgtgttttctcgtggtttgcttcaacatcgtaaacactgacaa 340  
DB 39540 CGCGGACCTGACACCGCAAGTCTTACAGCATCATCTTCGAAATGGTCAATCTGCTGG 39599  
QY 341 tcatttcacctcgtaattgttctgagcccaagattccattctcgtgttttct 400  
DB 39600 TTACCAATATTTTGGCCATGCTCG-----ACTCTCGCAGTGGCGGGCTCTTCTCT 39653  
QY 401 gggtaactcttgaaatggcgacttcaatcaattttagcaatgggcttatgaaactcgg 460  
DB 39654 GGACCACCATGGTGCATTTGTTTAACCTGTGTCACAGGCATCTACCAAGAACCA 39713  
QY 461 ttatggagttggtgagattttccgcaacactacattggtcgtctcttggattggaaca 520  
DB 39714 TATACGGAATAGTGGCATGCTTGGCCATCAAGTACACAGCGCGCTCGTCTGGCTCCA 39773  
QY 521 acatttgcgattgcgtgaacaacggttggaaatcggagtgacctattttctgaatgatg 580  
DB 39774 ACATAGCGGCTGCTTACACCGCCATGGCCCTTGATATGCGGTGAGATCTTCTCGTCCA 39833  
QY 581 agctaaacttgtgaactcgtctatttcggcctatctgttgggtgacctctgtgtgtg 640  
DB 39834 AG---AGGACTTCGGCCATCTACTTCTGTCGACCGCCATCTCTAGTCTGCTCTGCT 39890  
QY 641 caattgcaactttcttatacaaaaagaagatttaccactatcacccatcaaaaaggaa 700  
DB 39891 TCAGACAGTACTTTGGGCTGCCCTGAACAAGTCTTCGGGCACTACGAGACCATTAAGCC 39950  
QY 701 tggaaattcggaaaggcgaaacacgacagaccgtctccattcttctggaccacat 760  
DB 39951 GGAGCAGCGAAAGAAATCTGATTTCTAAGGCCACAGTGAACGTGCCTTACTTGGCAGATCT 40010

QY 761 tcacaaagtgttatgggcaactcttcaatgtttgggttctgtcttgcggttactctcaca 820  
DB 40011 TCAGAAAGCAGCGCCCAACTATTCAACATCTTCTCAGCTTTTTCGTCACGTTGTCGG 40070  
QY 821 tcttccctgttatgatgacggtttaccactcgtggagattccggtcttctactaaacaaat 880  
DB 40071 TTTTCCCGCGGATTCAGTCGACGCTCGACGCTCGACCCCAAAATTTTCGTA----- 40120  
QY 881 tgtctgaaacagatgaaatctacacttttgcctcaagtttctcctcgtcttctcaatttctg 940  
DB 40121 -----GTAGCCCGGACTACTTTTACACTTGTCTCTTCTTCGGGACCTTTTAAGCTCTCG 40175  
QY 941 ctgcatgtggtatccatagttgttccaaagattcactggtggcgcacaccccgcttacctcaat 1000  
DB 40176 CAATGTGGCAGTTTAACACCATGCTGGTGCATGCGCGGTGCAATGCGCGGTCTCTTGGG 40235  
QY 1001 ttgccataatcttgcgtgtcttttccattccattcttcttctctgcaacta-----tc 1054  
DB 40236 TGCACGTGGTCTTGGCGTCTGCGCTTCATCCCTTGTGTGTGTCATGTGCAACTAGCTTCCGC 40295  
QY 1055 gtgtcagacgctgtcttattcttcttcttcttcttcttcttcttcttcttcttcttctg 1114  
DB 40296 CGACCTCGGTGCGCTCATTTGGCGGTATTTCATGAGAACGATTGGGTGTAAGTGGGCATCG 40355  
QY 1115 gaattgccatgtcttttccatggtatccctcagcgtctggtgcaatgggatacactccaa 1174  
DB 40356 GCATCGCAATGGCTACAGCTCGGGTATCTCAGCTCTCTGGGCATGATGATGCGCCCGC 40415  
QY 1175 acgtctgcatctcactcactcaagatttgcgcctcagcttcttcttcttcttcttcttct 1234  
DB 40416 AGACGCTGCACACAAATACGACACCTGCGGAAATGTACGACAGTGCATGCTGATCA 40475  
QY 1235 ttggccttctcaacggtggtcgtgt 1258  
DB 40476 CGGGCATCTTCTCCGGAGTGTCT 40499

## RESULT 5

AE003614 259464 bp DNA INV 04-OCT-2000  
Drosophila melanogaster genomic scaffold 14200001386055 section 7  
of 63, complete sequence.  
ACCESSION AE003614 AE002690  
VERSION AE003614.2 GI:10728613  
KEYWORDS HTG  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

1 (bases 1 to 259464)  
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,  
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,  
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,  
Sutton,G.G., Wortman,J.R., Vandell,M.D., Zhang,Q., Chen,L.X.,  
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,  
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor  
Miklos,G.L., Abrell,J.F., Agbayani,A., An,H.J., Basu,A.,  
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Beeson,K.Y.,  
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,  
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkov,D.,  
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,  
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,  
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de  
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,  
Dunn,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,  
Dunson,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,  
Fleischmann,W., Foster,C., Gabrielian,A.E., Garr,N.S.,  
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,  
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,  
Hernandez,J.R., Houck,J., Hostin,B., Houston,K.A., Howland,T.J.,  
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,

Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacible, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wasserman, D.A., Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.

The genome sequence of *Drosophila melanogaster*  
Science 287 (5461), 2185-2195 (2000)  
20196006

2 (bases 1 to 259464)  
Adams, M.D., Celisnik, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.  
Direct Submission  
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA  
On Oct 9, 2000 this sequence version replaced gi:7297129.  
Location/Qualifiers  
1. 259464  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="2L"  
join(<538..1127,1190..1366,1421..>1745)  
/gene="CG9550"  
/product="CT27016"  
/db\_xref="FLYBASE:FBan0009550"  
/db\_xref="FLYBASE:FBgn0031826"  
/note="CG9550"  
/map="26D7-26D9"  
/db\_xref="FLYBASE:FBan0009550"  
/db\_xref="FLYBASE:FBgn0031826"  
join(538..1127,1190..1366,1421..1745)  
/gene="CG9550"  
/note="CG9550 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0009550"  
/db\_xref="FLYBASE:FBgn0031826"  
/protein\_id="AAF52397.1"  
/db\_xref="GI:7297130"  
/translation="MVSLSRTSKLTACVGVYTYTVLFIFFLLPMLPDPVTLQSLV  
SYSTSLGLDIRNLTLDAAGQPIRSMVLTVFRSGALTLDLNLAHQPGCVQHYAPLIAVE  
SRSTAKEHRLDELALALNCYDKSAEMIQWGRSPVFRFYCAQAKICRSYSOEA  
CWDEPTNAVKLYPFINMAYNLRLNLASLLERKDLNLMLLVDRPRTIYRMN  
NKWSDERDCQALCGDMVSDIQMVELTLQAYLQRFSLIRYEDLFLQDPDEIKQVED  
FYGLPLERNRIQKMHPRSGYFFQSSRWKLSQSAFSEWISMMVDDIGAVQDVCGQ  
2069..>3227  
/gene="CG9551"  
/product="CT27018"  
/db\_xref="FLYBASE:FBan0009551"  
/db\_xref="FLYBASE:FBgn0031827"  
/evidence="not\_experimental"  
<2069..>3227  
/map="26D7-26D9"  
/db\_xref="FLYBASE:FBan0009551"  
/db\_xref="FLYBASE:FBgn0031827"  
/evidence="not\_experimental"  
2280..3227  
/gene="CG9551"  
/note="CG9551 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0009551"

1  
/db\_xref="FLYBASE:FBan0031827"  
/evidence="not\_experimental"  
/protein\_id="AAF52398.1"  
/db\_xref="GI:7297131"  
/translation="MSTGIGSSSSSTEQOHLSEASDLGLGQDPDSNVDTGRT  
AKQLNGKRRRRMRARRRLLIGCVSSLCILLIATTHRRPLTTPFNQASGRDQ  
ASVGPSSNLARSFAFYDRFOQOQQSSQATVYNLTATIDVLSAQRSLAE  
NFEPGRGAERLTDMPTNGTPVRSVVTSWRSGSTFLGDLINSIPGNFYHPEPLD  
FGIKQIPDDQELAVONKLLNCDVADMDIYNFGKTHTYLFEHNTRLMDVCREFP  
RQWRPAFLTRFCFLFPQSMKTYRLRLAQAELLEOR"  
join(<27409..27443,28126..28308,28377..28578,  
28641..>28772)  
/gene="CG13765"  
/product="CT33249"  
/db\_xref="FLYBASE:FBan0013765"  
/db\_xref="FLYBASE:FBgn0031828"  
/evidence="not\_experimental"  
<27409..>28772  
/gene="CG13765"  
/map="26D9-26E1"  
/db\_xref="FLYBASE:FBan0013765"  
/db\_xref="FLYBASE:FBgn0031828"  
/evidence="not\_experimental"  
join(27409..27443,28126..28308,28377..28578,28641..28772)  
/gene="CG13765"  
/note="CG13765 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0013765"  
/db\_xref="FLYBASE:FBgn0031828"  
/evidence="not\_experimental"  
/protein\_id="AAF52399.1"  
/db\_xref="GI:7297132"  
/translation="WVMPQHVVISPLSSRVILVLRDPRCTMOSRRHRVWCGGNE  
EDPLVQDLRDDYKTAEVILLKYSRRTVRYEDLSLSPSEMQDILQFGLPDA  
VEEDLTHKYNIGVSYRSDRSAPFHMQDLKPEIKOIQDTCFAMDLGWVRI  
ENFNFTTQTFDPMWMPPT"  
complement(join(30215..30949,31026..31270,31333..33087,  
35664..35710,38587..38869))  
/gene="cli"  
/note="Nucleotide sequence of the Celera sequence differs  
from the published sequence for this transcript."  
/product="CT27024"  
/db\_xref="FLYBASE:FBan0009554"  
/db\_xref="FLYBASE:FBgn0000320"  
complement(<30215..>38869)  
/gene="cli"  
/note="CG9554"  
/map="26D9-26E1"  
/db\_xref="FLYBASE:FBan0009554"  
/db\_xref="FLYBASE:FBgn0000320"  
complement(join(30768..30949,31026..31270,31333..33087,  
35664..35710,38587..38658))  
/gene="cli"  
/note="cli gene product; Nucleotide sequence of the Celera  
sequence differs from the published sequence for this  
transcript"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0009554"  
/db\_xref="FLYBASE:FBgn0000320"  
/protein\_id="AAF52400.1"  
/db\_xref="GI:7297133"  
/translation="MVTLPYNYAARCGRLDKMIEPKVKRPKTDHTDTHERNLCNL  
SQOQQOQOQOQTHOQQOQOQOQSHSQSHSVLASGSSAGAGMGVGGGSGG  
GGVGGVGCSPGLGPQSQPLQPTIGLSLSGHYSNGNANPNWSSCSLATASSF  
AQASGSFTYQAGTSGGVGGGVGATVSHWTHDGTGSSAAVKSSRSPQV  
HASLDNGSVAGSNLYGCSSASNLDGGAVAVNSAAAAVADGKRDHYIYNSMQ  
YTPPPFYSGYTPYAATAARQAKMEPAAAAAYLTPSAAAGNNNSQLYSSPYAG  
YNNFGQDYGGYVNEQYGNYSYSPYSPSSSSASHGCHGFAAASNLSESPTD  
THSTTPVHTHSPSLPIPSPTCSGIGPLGNVSAASAAAALNSGSSVGTAGSGG  
VATSKTPTGKTGRARGRRHQPSPTSTRSTASDTGNSAEKPERVFWDLDTLIIFH  
TLLSGSANRYTKDHSLLMTAFRMEVFNMDTHFFTEIECDQVHIDVSDDN  
QDLSAYNATDGFHTNPPGAPPNCLPTGVGGVDMWRKLAFRYKIKDIYNSYRG  
NVGTILGPGKEAWLIQRSEIETVATDNWATLALCLSMISORENCVNVLTSTQLAPA



It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone F44D12 is at 1 in this sequence. The true right end of clone F44D12 is at 6268 in sequence 268299.

The true left end of clone T04B2 is at 31750 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence 268882.

The end of this sequence (31750..31853) overlaps with the start of sequence 268299.

For a graphical representation of this sequence and its analysis see: <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F44D12>.

## FEATURES

## source

## Location/Qualifiers

1..31853

/organism="Caenorhabditis elegans"

/db\_xref="taxon:6239"

/chromosome="IV"

/clone="F44D12"

join(3277..3413,3774..3951,3994..4239,4318..4603,

4647..4894,4945..5180,5225..5395,5449..5759,5805..6053,

6100..6802,6846..7020,7211..7306,7357..7425)

/gene="F44D12.1"

join(3277..3413,3774..3951,3994..4239,4318..4603,

4647..4894,4945..5180,5225..5395,5449..5759,5805..6053,

6100..6802,6846..7020,7211..7306,7357..7425)

/gene="F44D12.1"

/note="Similarity to Mouse Tight Junction protein ZO-1

(SW:ZO1\_MOUSE), contains similarity to Pfam domain:

PF00595 (PDZ domain (Also known as DHR or GLGF).),

Score=35.0, E-value=5.5e-07, N=2

cDNA EST yk16d12.5 comes from this gene

cDNA EST yk45f4.5 comes from this gene

cDNA EST yk35f8.5 comes from this gene

cDNA EST yk129c3.3 comes from this gene

cDNA EST yk129c3.5 comes from this gene

cDNA EST yk360b12.3 comes from this gene

cDNA EST yk360b12.5 comes from this gene

cDNA EST yk352d5.3 comes from this gene

cDNA EST yk352d5.5 comes from this gene

cDNA EST yk324a9.3 comes from this gene

cDNA EST yk324a9.5 comes from this gene

cDNA EST yk219a4.3 comes from this gene

cDNA EST yk219a4.5 comes from this gene

cDNA EST yk218f7.3 comes from this gene

cDNA EST yk218f7.5 comes from this gene

cDNA EST yk220g3.5 comes from this gene

cDNA EST yk252d9.5 comes from this gene

cDNA EST yk357f4.5 comes from this gene

cDNA EST yk388f10.5 comes from this gene"

/codon\_start=1

/protein\_id="CAA92607.1"

/db\_xref="gi:3871158"

/db\_xref="SPTREMBL:Q20398"

/translation="MEKEIREVEAKMEKVRKASERDDIGKKIDFEPEKIDSLQIQV

HRSTYIEEANKRLRLOYEALQDYETEAOERSIVLEENRDEERDRLOKEIEK

MKEDEEVAALKLENTFRLLKNWETAQANAEAAIARLTAEANGARITFERD

EALRKCOQTGDAIVDINWTHSVQINLPNPNLGMIGGGRTDGSIVHGPIYVQRI

AHGSFDFNLKLDHIMVNDISVTDMDERSVGMLSNCHHILVIRRSNCKIISDV

CLPNYGLSELNGVFINSCPEGAASRGLAQGVVHVHMTPVYDAKHAEMLIK

SREPLVIGILASTRGDNGDKHRTQIFSRWFSRNGSGDKERTVVAKANIDRSNDQV

LLRGLSLMPQASPMSPLYRYSRAPTYSSSIDHTKMLDALDKRFNAKSTAHSE

ISSSTAPTAALWSPSSVNEKDLVPACNIEGVVYVPKSTIESTPTSPVTRLRQNS

EOKRTMSIISRHWGDRGSSNTTSDARPSVPHFTPTSTIMEGRPHRSAYISP

SHPPVPTIRDDSSMSSNSIRLPTSPFSNOYPNNTSLTSGGVPVPRHCSISKD

GSDFSTIGTAMSHSIAEGRNKKRYHLPRGNHSESLVNRSSQRLVEVPRSDVLQK

GSGAIGILAEKISIGDLDGLILSDGNCVNTLECAMNTLSADNSELCTGLVQDC

GDRLNRLPLGADGSFFLRVINDRSMENKDELDKCGDVFEVDKTMLGKTCRWPAWK

VDKRGROHECAISPTSYVQAIRANRYANFPKAYEWEKLDTKVKRPRLLFGAVV

EPFLQMLVDESKFYSIVARELSALLDKVDLIDKQNDVDYDLVHVVSTAHL

IMDTAQGLHCLVQDQSAIDRLKRCRMFFILVKIRFKSVKQLKDVBHIGCEKISSK

EAKQLIEKDLKIEKDLGDSVTLVVPSHNNVFMTHAVLQLKKLIIEDEQKKIVWVQRK

## VDEE"

complement(join(7556..7673,7717..7832,8096..8296,

8340..8510,8562..8718,8944..9111,9541..9638))

/gene="F44D12.2"

complement(join(7556..7673,7717..7832,8096..8296,

8340..8510,8562..8718,8944..9111,9541..9638))

/gene="F44D12.2"

/note="cDNA EST yk169e12.3 comes from this gene

cDNA EST yk367d4.3 comes from this gene

cDNA EST yk367d4.5 comes from this gene

cDNA EST yk553h9.3 comes from this gene

cDNA EST yk553h9.5 comes from this gene"

/codon\_start=1

/protein\_id="CAA92598.2"

/db\_xref="gi:5824507"

/db\_xref="SPTREMBL:Q20392"

/translation="MOTILQILIVLTLFHEVNTLVTCRCRFGNECNFSRSTRCRGD

FCYHAQFYGNEMMPTIQKGCVIGEYVPCGRVNHGHNVCFCSDADYCNANTSLAD

TATITLPTQTCQPEKVNMPKRWTKPCAANYCTFTEAKTQTEVNSANTWTSCDNK

ENEFDFPPTLVFNFPYGVTCVWLNGQPDTHACVGSDSLDTLTFDVTATTTECHVD

YFNRLPYVKSGSSCIGQFCFISATSRGVEFGCVNSVTVEGATPLKIGVTRAYTGLE

QNICQSYCNADLKSNELSWPPELYLYRNISNLRNFVFIIDSRSSSIFLAIPVIF

LINYCLFHF

join(12326..12380,12427..12481,12574..12895)

/gene="F44D12.9"

join(12326..12380,12427..12481,12574..12895)

/gene="F44D12.9"

/codon\_start=1

/protein\_id="CAA92605.1"

/db\_xref="gi:3871156"

/db\_xref="SPTREMBL:Q20396"

/translation="MTKMQQEEGSPLSVFOYLHVEYNICRKGQOTTLKMNELAA

ENSSVTGNSQIGLVTPRDTWRKESPTDRWLVYLIFTMHGMLMSNMNFITIAPO

YVHDVFNFTNYQDSEMSIIGVTSQIPNVGIMLTIVMVV"

complement(15660..15989)

/gene="F44D12.3"

complement(15660..15989)

/gene="F44D12.3"

/note="contains similarity to Pfam domain: PF00635 (MSP

(Major sperm protein) domain), Score=111.7,

E-value=4.7e-30, N=1

cDNA EST yk57e8.3 comes from this gene; cDNA EST yk102a1.3

comes from this gene

cDNA EST yk57e8.5 comes from this gene; cDNA EST yk102a1.5

comes from this gene

cDNA EST yk122a5.5 comes from this gene; cDNA EST

yk149d3.5 comes from this gene

cDNA EST yk151a2.5 comes from this gene; cDNA EST

yk164g1.5 comes from this gene

cDNA EST yk170h1.3 comes from this gene; cDNA EST

yk179d6.5 comes from this gene; cDNA EST

yk149d3.3 comes from this gene; cDNA EST

cDNA EST yk151a2.3 comes from this gene; cDNA EST

yk164g1.3 comes from this gene

cDNA EST yk170h1.3 comes from this gene; cDNA EST

yk179d6.3 comes from this gene

cDNA EST yk189f1.3 comes from this gene; cDNA EST

yk19a10.3 comes from this gene

cDNA EST yk92f10.3 comes from this gene; cDNA EST

yk15a11.5 comes from this gene

cDNA EST yk164f1.5 comes from this gene; cDNA EST

yk378f3.3 comes from this gene

cDNA EST yk378f3.5 comes from this gene; cDNA EST

yk431f3.3 comes from this gene

cDNA EST yk431f3.5 comes from this gene; cDNA EST

yk365g9.3 comes from this gene

cDNA EST yk365g9.5 comes from this gene; cDNA EST

yk311b6.3 comes from this gene

cDNA EST yk311b6.5 comes from this gene; cDNA EST

yk299b2.3 comes from this gene

cDNA EST yk299b2.5 comes from this gene; cDNA EST

yk309f7.3 comes from this gene



```
cDNA EST yk121b5.3 comes from this gene
cDNA EST yk121b5.5 comes from this gene
cDNA EST yk262b6.3 comes from this gene
cDNA EST yk262b6.3 comes from this gene
/codon_start=1
/protein_id="CAB62794.1"
/db_xref="GI:6562352"
/translation="MONNVDSIIIDVLSASTHEKPLCKVITERVVLKLLDLALGVFK
AOKPWEVNAIPKVGDIHGQDPDLRLFRHGGWPTANYLFLGQDVDRGRFSIETIV
LLAYKYPGCFNFFLLRGNHCEFNKTYGYECOKRQSVRYAAQDFVFNWLPIT
GLIATKILMGHGLSPLMKFTLDTLRKIERPTGEGVLADLLWADPISGLSGFMN
NORGACGFGDSVLNLCSEFQDLVRAHQVQDGYEFFAGKLVITIFSAPHYCCQF
DNCAAFMSCDKLQCSFEILRPTTGRLEIREKPLKDET"
33294..33521
/gene="C47A4.5"
33294..33521
/gene="C47A4.5"
/codon_start=1
/protein_id="CAB62795.1"
/db_xref="GI:6562353"
/translation="MADKSARVPVDAIGNHNKNTDLVDIDDELFCCKPKPKTAEPKPA
PPPPAPAPAAPSTGKQYKKSSTYKSYAK"
BASE COUNT 13436 a 6918 c 6544 g 13766 t
ORIGIN

Query Match 5.1%; Score 71.2; DB 3; Length 40664;
Best Local Similarity 53.6%; Pred. No. 1.7e-07;
Matches 148; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 390 cgggttttttcggtaaccttgggaatgagcgaacttaacatttagcaatgggctata 449
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17758 CTGGTCTACATTGCTCTATTATCATAGTAATGCAATGAATGATCCACGGCTCTCA 17817

QY 450 tgaataactcggtttatggaggttgggtggaatttcgcgcacactacattggcgctcttt 509
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17818 CCAAAACTCATTTTTCGGAATGGCTGCAGATTTCCTCCGCAAAATACTCGAACGCTGTGGT 17877

QY 510 gattggaacaacatttcgggattgctgataacaggttggtaaaatggagtgacctattt 569
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17878 TATTGGAACAAATATTGTGGGAACATTCTACTAGTGTCTTGGCAATGTGCGAACCTTTGGC 17937

QY 570 tctgaatgatgacgttaacttggcaatcgtctacttcgcatatcgcttggtgatcct 629
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17938 ATTCTACACACAGCTGAACACTGTGCACTTATTATTATTGGAATATCACTCTTATTCT 17997

QY 630 tctggtgtgcaattgcaactttttttatcacaa 665
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17998 GTTTGTTGCTGTGTCATGCTGGTGTCTGCAAGAA 18033

RESULT 8
CEZK809/c 25560 bp DNA INV 20-JUN-2001
LOCUS Caenorhabditis elegans cosmid zK809, complete sequence.
DEFINITION z68303
ACCESSION z68303
VERSION z68303.1 GI:1130662
KEYWORDS HTG; Hydrophobic nucleolar protein like; Peroxisome assembly factor
like; Protein-tyrosine phosphatase.
SOURCE Caenorhabditis elegans.
ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 25560)
REFERENCE none.
AUTHORS Genome sequence of the nematode C. elegans: a platform for
TITLE investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
REMARK The C.elegans Sequencing Consortium.
REFERENCE 2 (bases 1 to 25560)
```

AUTHORS  
TITLE  
JOURNAL

## COMMENT

FEATURES  
source

## gene

## CDS

## gene

## CDS

Dobson, R.  
Direct Submission  
Submitted (20-DEC-1995) Nematode Sequencing Project, Sanger Centre,  
Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,  
Washington University, St. Louis, MO 63110, USA. E-mail:  
jes@sanger.ac.uk or rw@nematode.wustl.edu  
Coding sequences below are predicted from computer analysis, using  
predictions from GeneFinder (P. Green, U. Washington), and other  
available information.  
Current sequence finishing criteria for the C. elegans genome  
sequencing consortium are that all bases are either sequenced  
unambiguously on both strands, or on a single strand with both a  
dye primer and dye terminator reaction, from distinct subclones.  
Exceptions are indicated by an explicit note.  
IMPORTANT: This sequence is not the entire insert of clone zK809.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring submissions.  
The true left end of clone zK809 is at 1 in this sequence. The true  
right end of clone zK809 is at 15121 in  
sequence z68302.  
The true left end of clone zK792 is at 25461 in this sequence. The  
start of this sequence (1..113) overlaps with the end of sequence  
z73427.  
The end of this sequence (25461..25560) overlaps with the start of  
sequence z68302.  
For a graphical representation of this sequence and its analysis  
see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=zK809)  
name=zK809.

Location/Qualifiers  
1..25560  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/chromosome="IV"  
/clone="zK809"  
join(1372..1376,1431..2001,2054..2551,2599..2803,  
2847..2854)  
/gene="zK809.1"  
join(1372..1376,1431..2001,2054..2551,2599..2803,  
2847..2854)  
/gene="zK809.1"  
/note="contains smilarity to Pfam domain: PF00102  
(protein-tyrosine phosphatase), Score=35.2,  
E-value=2.3e-13, N=1  
cDNA EST yk365g6.3 comes from this gene"  
/codon\_start=1  
/protein\_id="CAA92641.2"  
/db\_xref="GI:5824906"  
/translation="MPNKGVGKKSPSTKSSNNKKRRREPESRYEEKTAVHGGTKTYE  
EKTVVHGGTKNVVAKVPVSKASVIVDKKASTTETVPTTVKKSMLSISTAYDRTVTE  
EXEKPKKKEKESGEDASIKFELRENOSEGVRDTEASCNRCWKWQSHSRN  
RHTDYKYNHRIIVOMCKDDYINGSKINVPHTPSIYLVLPKMSVDVAEFRVW  
FHEQCQTVHIIARPELTNPAIDKLCFQSGAWLYANGFFVQVTRKVEKNKADMFV  
VELLPQCSNAVCMCSYLHTYKPFDPFGAQIRAAHQIAKNENKNSPTVIASNG  
SGRNAALLTAVVEDQLTGRKPKISEIVRTIREQRQSDVSIQVVSLYMATNWLK  
TKLPSEHVVSFKIKFNEKATRSPLGSMKTPTEEL"  
join(3252..3612,3660..3835,4567..4606,5066..5406)  
/gene="zK809.7"  
join(3252..3612,3660..3835,4567..4606,5066..5406)  
/gene="zK809.7"  
/note="Similarity to Rat peroxisome assembly factor-1  
(SW:PAFL1\_RAT), contains similarity to Pfam domain:  
(zinc finger, C3HC4 type (RING finger)), Score=16.2,  
E-value=0.00035, N=1  
cDNA EST yk360a11.5 comes from this gene"  
/codon\_start=1  
/protein\_id="CAA92640.1"  
/db\_xref="GI:3881798"  
/db\_xref="SPTREMBL:Q23601"  
/translation="MIHFYHOKRNILNNTLHCTLPHYHYFSVFLSDTSFYLLF  
SFVSITFLYFKMACRQDQLDAHVLDSEIRIAQDSVDDVSNOMPLWVARIFEKLR  
PELKVTLEAVLWTHRFSSRGASQEQMDIAYAGYTPKLTITSHFMVEVIVPILTRMSE

LSGRIDISRWYSKIEAIYELGSLHFLYFLRKSASINYDTQNRELMMHAFRDVILLTYP  
FIEKARQIRIVKQKLDKRFSSIGSSDIECVKDCPSVIMVQKCGHVACYTCIATS  
RNWTCLSCSNAEMFLAKLNPLIE"  
complement(join(10131..10337,10438..10717,10774..10870,  
10916..11035,11085..11127))  
/gene="ZK809.2"  
complement(join(10131..10337,10438..10717,10774..10870,  
10916..11035,11085..11127))  
/gene="ZK809.2"  
/note="contains similarity to Pfam domain: PF01553  
(Acyltransferase), Score=144.3, E-value=7e-40, N=1  
cdna EST yk576b1.3 comes from this gene  
cdna EST yk576b1.5 comes from this gene  
/codon\_start=1  
/protein\_id="CAA92638.1"  
/db\_xref="GI:3881796"  
/db\_xref="SPTREMBL:O23598"  
/translation="MSLVTSVSKLMFLGGSKNLICHNKETFFVKILENPNPLITVSNH  
RSNIDPLMWCILKRFWRKYKDRNRYTLAAHICFTKQPHTFMSLGRVPCVGRGEG  
VYQKGMDFCDMLNDKRWVHIFPEGKVTTLSEPLRFKMGIGRLVMDAKTDPVILPVW  
CKMEKVMPTQPPYYPKFGNTVTVHIGEPFLSDLKTVLSKSLTTEQMRKIITDEVQ  
TRMVLGEKVGDLPGKSSLEILRKNPIEV"  
join(11625..11731,11785..11887,11940..12056,12122..12232,  
12286..12419,12469..12529)  
/gene="ZK809.3"  
join(11625..11731,11785..11887,11940..12056,12122..12232,  
12286..12419,12469..12529)  
/gene="ZK809.3"  
/note="predicted using Genefinder  
cdna EST yk78b3.3 comes from this gene; cdna EST yk78g8.3  
comes from this gene  
cdna EST yk105h6.3 comes from this gene; cdna EST yk78b3.5  
comes from this gene  
cdna EST yk78g8.5 comes from this gene; cdna EST yk105h6.5  
comes from this gene  
cdna EST yk636e8.3 comes from this gene; cdna EST  
yk636e8.5 comes from this gene  
cdna EST yk437a12.3 comes from this gene; cdna EST  
yk437a12.5 comes from this gene  
cdna EST yk459f11.3 comes from this gene; cdna EST  
yk459f11.5 comes from this gene  
cdna EST yk456e1.3 comes from this gene; cdna EST  
yk456e1.5 comes from this gene  
cdna EST yk435c7.3 comes from this gene; cdna EST  
yk435c7.5 comes from this gene  
cdna EST yk435b11.3 comes from this gene; cdna EST  
yk435b11.5 comes from this gene  
cdna EST yk417d5.3 comes from this gene; cdna EST  
yk417d5.5 comes from this gene  
cdna EST yk391b6.3 comes from this gene; cdna EST  
yk391b6.5 comes from this gene  
cdna EST yk197b7.3 comes from this gene; cdna EST  
yk197b7.5 comes from this gene  
cdna EST yk199b5.3 comes from this gene; cdna EST  
yk199b5.5 comes from this gene  
cdna EST yk538d11.3 comes from this gene; cdna EST  
yk561e2.3 comes from this gene  
cdna EST yk567a11.3 comes from this gene; cdna EST  
yk586e8.3 comes from this gene  
cdna EST yk597b7.3 comes from this gene; cdna EST  
yk616f10.3 comes from this gene  
cdna EST yk620g3.3 comes from this gene; cdna EST  
yk623f9.3 comes from this gene  
cdna EST yk650c6.3 comes from this gene; cdna EST  
yk655c5.3 comes from this gene  
cdna EST yk665g12.3 comes from this gene; cdna EST  
yk538d11.5 comes from this gene  
cdna EST yk561e2.5 comes from this gene; cdna EST  
yk567a11.5 comes from this gene  
cdna EST yk586g8.5 comes from this gene; cdna EST  
yk597b7.5 comes from this gene  
cdna EST yk616f10.5 comes from this gene; cdna EST  
yk616h5.5 comes from this gene

cdna EST yk620g3.5 comes from this gene; cdna EST  
yk623f9.5 comes from this gene  
cdna EST yk650c6.5 comes from this gene; cdna EST  
yk655c5.5 comes from this gene  
cdna EST yk665g12.5 comes from this gene  
/codon\_start=1  
/protein\_id="CAA92637.1"  
/db\_xref="GI:3881795"  
/db\_xref="SPTREMBL:O23597"  
/translation="MGNQYKPKERTGNSPYTRSGVPPALVKTGDKPLHRVNDMSLE  
LHADERVRAAGLSFADREWRKWHQDHLHADEPVVDVHRQLNPIRTAYRLPWDR  
FYHLKPTFGVYGTATYAPKLMFAFVVQYAYYWKYEVKDWTHLRLGLESMPQK  
EVIASAPAIIEKPEGLMDKGLANPSSKDDYTPFTFNKRTAYLDVGETRPW"  
complement(join(12918..13232,13332..13978,14028..14122,  
14201..14367,14519..14680))  
/gene="ZK809.4"  
complement(join(12918..13232,13332..13978,14028..14122,  
14201..14367,14519..14680))  
/gene="ZK809.4"  
/note="Similarity to Human 36k hydrophobic nucleolar  
protein (PIR Acc. No. JC4196), contains similarity to Pfam  
domain: PF01733 (Nucleoside transporter), Score=541.4,  
E-value=2e-159, N=1  
cdna EST yk77f11.3 comes from this gene  
cdna EST yk77f11.5 comes from this gene  
cdna EST yk373c3.3 comes from this gene  
cdna EST yk373c3.5 comes from this gene  
cdna EST yk328c5.3 comes from this gene  
cdna EST yk328c5.5 comes from this gene  
/codon\_start=1  
/protein\_id="CAA92642.1"  
/db\_xref="GI:3881800"  
/db\_xref="SPTREMBL:O23599"  
/translation="MSSAVELQPLNKTKKVEESEPEDKGNLYFIILLHGIGTLMP  
WNMLITISDYDFESYKMLANSTIDMTGVYGTPTVYSSNFSQQTASQVPLLNL  
LNIFIVVKGGLASRITVGLSIVAVCVITTMFIYVETSTWLTGFTTITITIIVLNGA

	Query Match	5.1%	Score 70.8;	DB 3;	Length 25560;
	Best Local Similarity	52.0%;	Pred. No. 2e-07;		
	Matches 159;	Conservative 0;	Mismatches 147;	Indels 0;	Gaps 0;
Qy	978	gccagacacccggttacctcaaatgtgccaatcttgcgtagctctcttccattccattctt	1037		
Db	13233	GCCTCGCGCGCAACAGCTCTGGATTCCAGTCTACCTCGCTCTCTATATTCATTCTT	13174		
Qy	1038	ctctcttgcacatctgtgtccagacgcgtcttctctcttcttgcgtagctctactga	1097		
Db	13173	CATCTTCGCAACATATCTCCCTGAGACCCGTTCACTTCCCGTCTTTTTCGAAATCCACCTG	13114		
Qy	1098	cattttgtgattggtggaattgccaatgtcttttccatcgataccatcagcgtctggc	1157		
Db	13113	GCTTTTCGTCATTATCGCGCGCTCGATGAGCTTTGGAAGTGGATATTTCTCGGGACTCGC	13054		
Qy	1158	aatgggatacactccaaacgctgctgcaatctcactactcaagattgacgctcagcttcc	1217		
Db	13053	CATGATGTACACCTCGAAGACTGTTGATCCATCGAAGCTCAGGTCGCTGGAATGATGCG	12994		
Qy	1218	cggttgcaactcttatggttgcccttctcaccggtggccctgtggccggttgattatgagca	1277		
Db	12993	CGGAATCTTCCTCATCTCTCGAATTGCTCTGGTCTGATCTTACATGGTCATCAAGAT	12934		
Qy	1278	cttcgt 1283			
Db	12933	GGTCGT 12928			

RESULT 9  
LOCUS I66494 7218 bp DNA PAT 28-DEC-1997  
DEFINITION Sequence 14 from patent US 5670367.  
ACCESSION I66494  
VERSION I66494.1 GI:2724471  
KEYWORDS



jess@sanger.ac.uk or rw@ematode.wustl.edu  
 On Sep 24, 1998 this sequence version replaced gi:1515146.  
 Coding sequences below are predicted from computer analysis, using  
 predictions from Genefinder (P. Green, U. Washington), and other  
 available information.  
 Current sequence finishing criteria for the C. elegans genome  
 sequencing consortium are that all bases are either sequenced  
 unambiguously on both strands, or on a single strand with both a  
 dye primer and dye terminator reaction, from distinct subclones.  
 Exceptions are indicated by an explicit note.  
 IMPORTANT: This sequence is not the entire insert of clone K09A9.  
 It may be shorter because we only sequence overlapping sections  
 once, or longer because we arrange for a small overlap between  
 neighbouring submissions.  
 The true left end of clone K09A9 is at 35278 in  
 sequence 279596.  
 The true right end of clone K09A9 is at 34017 in this sequence. The  
 true left end of clone K09E9 is at 24804 in this sequence. The true  
 right end of clone C02C6 is at 104 in this sequence. The start of  
 this sequence (1..104) overlaps with the end of sequence 279596.  
 The end of this sequence (33914..34017) overlaps with the start of  
 sequence 279602.  
 For a graphical representation of this sequence and its analysis  
 see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=K09A9)  
 name=K09A9.

FEATURES	Location/Qualifiers
source	1..34017
	/organism="Caenorhabditis elegans"
	/db_xref="taxon:6239"
	/chromosomes="X"
gene	/clone="K09A9"
	complement(join(9270..9399,10952..11597,11646..11868,12089..12263,12317..12450,12502..12642))
	/genes="K09A9..5"
CDS	complement(join(9270..9399,10952..11597,11646..11868,12089..12263,12317..12450,12502..12642))
	/genes="K09A9..5"
	/note="Predicted using Genefinder
	Similarity to Bovine NADH-ubiquinone oxidoreductase 49 KD subunit (SW:NUCM_BOVIN), contains similarity to Pfam domain: PF00346 (Respiratory-chain NADH dehydrogenase, 49 KD subunit). Score=595.8, E-value=6.4e-215, N=1
	CNA EST EMBL:Z14772 comes from this gene
	CNA EST EMBL:M88875 comes from this gene
	CNA EST EMBL:M89294 comes from this gene
	CNA EST EMBL:T01723 comes from this gene
	CNA EST Yk81b6.3 comes from this gene
	CNA EST Yk81b6.5 comes from this gene
	CNA EST Yk68h5.3 comes from this gene
	CNA EST Yk68h5.5 comes from this gene
	CNA EST Yk70h1.5 comes from this gene
	CNA EST Yk70h1.3 comes from this gene
	CNA EST Yk501g3.3 comes from this gene
	CNA EST Yk485g3.3 comes from this gene
	CNA EST Yk485g3.5 comes from this gene
	CNA EST Yk454f8.3 comes from this gene
	CNA EST Yk454f8.5 comes from this gene
	CNA EST Yk424c6.3 comes from this gene
	CNA EST Yk424c6.5 comes from this gene
	CNA EST Yk418a5.3 comes from this gene
	CNA EST Yk418a5.5 comes from this gene
	CNA EST Yk229c10.3 comes from this gene
	CNA EST Yk229c10.5 comes from this gene
	CNA EST Yk409d1.5 comes from this gene
	CNA EST Yk530c9.3 comes from this gene
	CNA EST Yk532b10.3 comes from this gene
	CNA EST Yk596h9.3 comes from this gene
	CNA EST Yk652g6.3 comes from this gene
	CNA EST Yk680a8.3 comes from this gene
	CNA EST Yk530c9.5 comes from this gene
	CNA EST Yk562f5.5 comes from this gene
	CNA EST Yk675a3.5 comes from this gene"
	/codon_start=1

Query Match 4.9%; Score 68; DB 6; Length 7218;  
 Best Local Similarity 4.2%; Pred. No. 9e-07;  
 Matches 17; Conservative 234; Mismatches 149; Indels 0; Gaps 0;

Qy	907	ttgtcacaagtttcctcgcttcaattgttcgtcgatggtgacatagttgcttc	966
Db	1060	TTGCGATYYY	1119
Qy	967	aagattcactgcccagaccctgttacctcaaatggccataatctgctgctcttc	1026
Db	1120	YYY	1179
Qy	1027	attcattcttcttctgcaacatcgtgtccagacgctgtattctcttctt	1086
Db	1180	YYY	1239
Qy	1087	gagctactgacatttggattggtgggaattgccatgctttttcacatgatactc	1146
Db	1240	YYY	1299
Qy	1147	agcgtctgcaatggatgatacactccaaacgtggtccactcactactcaagattgcc	1206
Db	1300	YYY	1359
Qy	1207	gctcagcttcctggttcgactcttattggtgaccttcacccgtggtggtccgtt	1266
Db	1360	YYY	1419
Qy	1267	gtattgagcacttcgtggacaagcgaagtatctataaa	1306
Db	1420	YYYYYYYYYYYYYGTTACCAAAATCTCTATCTCTTAA	1459

RESULT 10  
 CEK09A9  
 LOCUS 34017 bp DNA INV 20-JUN-2001  
 DEFINITION Caenorhabditis elegans cosmid K09A9, complete sequence.  
 ACCESSION 279601  
 VERSION 279601.1 GI:3646501  
 KEYWORDS HM6; Aspartyl beta hydroxylase like; NADH-ubiquinol oxidoreductase; protein kinase; RAS-related protein; Ubiquitin carboxyl-terminal hydrolase.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 34017)  
 none.  
 Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium  
 Science 282 (5396), 2012-2018 (1998)  
 99063613  
 MEDLINE The C.elegans Sequencing Consortium.  
 REFERENCE 2 (bases 1 to 34017)  
 AUTHORS Swinburne,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-AUG-1996) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:



N-acetylglucosamine-1-phosphate transferase; protein kinase; R-CG7 repeat; R-FA3 repeat; repl1; rep20; rifin; RNA-binding protein; stevor; T-complex protein 1 epsilon subunit; telomere; var. malaria parasite P. falciparum.

## SOURCE

ORGANISM

Plasmodium falciparum  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
1 (bases 1 to 253305)  
Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C. M., Craig, A., Davies, R. M., Devlin, K., Feltwell, T., Gentles, S., Gwilliam, R., Hamlin, N., Harris, D., Holroyd, S., Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Kyes, S., McLean, J., Moule, S., Mungall, K., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutter, S., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Whitehead, S., Woodward, J. R., Newbold, C. and Barrell, B. G.

## TITLE

The complete nucleotide sequence of chromosome 3 of Plasmodium

## JOURNAL

falciparum

## MEDLINE

Nature 400 (6744), 532-538 (1999)

## REFERENCE

99376085

## AUTHORS

2 (bases 1 to 253305)

## JOURNAL

Oliver, K., Bowman, S., Harris, D., Lawson, D., Quail, M. and Barrell, B.

## REFERENCE

3 (bases 1 to 253305)

## AUTHORS

Lawson, D., Bowman, S. and Barrell, B.

## TITLE

Direct Submission

## JOURNAL

Submitted (17-DEC-1998) P. falciparum Genome Sequencing Consortium,

The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge

CB10 1SA, UK

## COMMENT

On May 14, 2001 this sequence version replaced gi:2982540

gi:2982541 gi:2982544 gi:2894453 gi:2894464 gi:2982551 gi:2982556

gi:2982558 gi:2982563 gi:2982564 gi:2982566 gi:2894491 gi:2982569

gi:2695974 gi:2894502 gi:2673766 gi:2894496 gi:2982577 gi:2894588

gi:2894589 gi:2982538 gi:2982539 gi:2894380 gi:4725992.

For more information about this sequence or the Malaria Project,

see [http://www.sanger.ac.uk/Projects/P\\_falciparum](http://www.sanger.ac.uk/Projects/P_falciparum).

## FEATURES

## source

location/Qualifiers

1. 253305

/organism="Plasmodium falciparum"

/strain="3D7"

/db\_xref="taxon:5833"

/chromosome="3"

/clone="MAL3P7"

458..4060

/gene="PFC0860w, MAL3P7.1"

458..4060

/gene="PFC0860w, MAL3P7.1"

/note="PFC0860w (MAL3P7.1), Kinesin-related protein, len:

1200 aa; similarity: to Kinesin-related protein.

D.melanogaster kinesin-like protein 57A (TR:P91945) BLAST

Score: 664, sum p(2) = 1.9e-67; 28% identity in 707 aa

overlap, predicted using hexExon, pfam: match to PF00225

kinesin, kinesin motor domain Score 311.54"

/codon\_start=1

/product="kinesin-related protein"

/protein\_id="CAB39023.1"

/db\_xref="GI:4493964"

/db\_xref="SPTREMBL:O97277"

/translation="MRKTKNLDPFLNNYYKNEENPLIICSDERKTKIYSTNH

HTNNILYNHTLKLICNTENNAKILHSKNSLQKNILMPSYQKRGTHIRETIK

VYPINGEPSTVENTINGEHFINGQYDALKNSLNNDHOHNNMNNISNNKLFLV

NNNNNNQDVENNNPITSSYTKNNNNLNTLNAYSNVKAIVRIKPIGESENIYS

IFNKNYLLIEKNEKCYLLISQKKQSTYVDSVFDVNAQVEEFFQTAAPLIPHVK

GINTVFAYGATGSGKTYTMDLDDKNQNGIVQLSLELFTIINERKCRNIKVLMSFLEV

YNTIRDLGKGNKTLVEQEDVAEVKSNICELEVNNYEQAMLLINEGVKRNKMSFT

RANKVSRSHALIQIYYNEILDNNMTISYKALCFVDLAGSERASATSNKGRFKE

GSYINQSLLALANCISLANRNSIKVRVYRDSKLTHLLKNSLEGNCVLYMIANIP

SRTSPQSNLTAFAPARNIKCATQYTDNDKESDIEKILKKNENLQKEDYDILGKY

TLNKEFFPIINQLYKKQISCYLLKNTISDNNSSMELKODITMYDQLYKMKSDYVR

KKVDSLQYQEEKQFLNLFDTFLKLNANVINSKVDNNKSLLEEMIFFKNEKN

VNENFLVNEKVDKNVINGVNDENVVDKNVINGVNDENVVDENVVDENVVDENV

VDNVVDENVVDENVVDENVVDENVVDENVVDKNVSNENIIVSEKHKVGLS

AEKSESHNNKNDIEDNDKDTIKDIHNNNNNSDNDNDEYQSANSPVESDIVKKEK

## misc\_feature

/gene="PFC0860w, MAL3P7.1"

/note="potential splice donor at 3' end of PFC0860w may indicate splicing"

4057..4062

## misc\_feature

/note="potential splice donor at 3' end of PFC0860w may indicate splicing"

join(5457..5577,5689..5787,5923..6077,6189..6743)

/gene="PFC0865w, MAL3P7.2"

join(5457..5577,5689..5787,5923..6077,6189..6743)

/gene="PFC0865w, MAL3P7.2"

/note="PFC0865w (MAL3P7.2), RNA-binding protein, len: 309

aa; Similarity eg to C.elegans RNA-binding protein

(TR:Q18318) BLAST Score: 378, sum p(1) = 3.8e-35; 37%

identity in 297 aa overlap, predicted using hexExon, pfam:

match to PF00076 rrm, RNA recognition motif. (a.k.a. RRM,

RBD, or RNP domain) Score 58.01"

/codon\_start=1

/product="putative RNA-binding protein"

/protein\_id="CAB39067.1"

/db\_xref="GI:4494008"

/db\_xref="SPTREMBL:O97318"

/translation="MGLPDKIRNIEKLEAEKLNIGNNDSSWHDOYRDSSYIYIGNLD

NRLPGDIVFVSQYGEIDVNLVRDNETGSKGYCFLSYADORSITLADVNFNGYKL

LERLPVHLIINLRPKYLDKADNEYKPAEGGIGVTVNSEIKLSKVFQKIK

NKSEDDKKLLDEDELWALNFKSIIKDIISPIGHDEKSRHNEGKEEDEDDED

DEDEDDESDVDIKYKRHKRKSITTKYDKDKRKHRSRHRHNSHREK

EKKDKSHRRKHSHYDKYSRSRSYSTSSYDTR"

join(7801..7928,8084..8234,8392..8583)

/gene="PFC0870w, MAL3P7.3"

join(7801..7928,8084..8234,8392..8583)

/gene="PFC0870w, MAL3P7.3"

/note="PFC0870w (MAL3P7.3), putative elongation factor,

len: 181 aa; Similarity to P. falciparum elongation factor

1-beta, (AAF27524) (276 aa), fasta scores: opt: 366, E(1):

1.1e-15, 53.2% identity in 141 aa overlap, revised:

shortened exon 2, Pfam: match to PF00736 EF1B0, EF-1

guanine nucleotide exchange domain Score 61.10"

/codon\_start=1

/product="putative elongation factor"

/protein\_id="CAB39068.2"

/db\_xref="GI:8052274"

/db\_xref="SPTREMBL:O97319"

/translation="MANTYDELYPLSYIYLLONEGNTSKIDQANTKKPKKEVINKS

SLIIDIKPYGENTDLDVILKLVNITMEGLTWGKAHKKTPEAFGLFKIQVSCVIVDDL

VNTDELITENLGLDNEQLQKKQMDDEENYDEDDITGGLVQSAEIIISPNKL"

8235..8240

/gene="PFC0870w, MAL3P7.3"

/note="revised splice donor sequence for exon 2 of

PFC0870w"

8580..8585

/note="potential splice donor sequence may indicate

further splicing of PFC0870w"

join(11431..20416,20578..20690)

/gene="PFC0875w, MAL3P7.4"

join(11431..20416,20578..20690)

/note="PFC0875w (MAL3P7.4), hypothetical protein, len:

3032 aa, revised: added 3' exon, possibly spliced at 5'

end"

/codon\_start=1

/product="hypothetical protein, PFC0875w"

/protein\_id="CAB39024.2"

/db\_xref="GI:8052275"

/db\_xref="SPTREMBL:O97278"

/translation="MKDPLSYAENNDLNFILKLYNSEEDCLNKLKVAIELKSEDFKNVN

NPLHNDLILNTHENDKNLLNLTNINLKSEFPNIDLYKDVITENEFENKLNKI

NPILKLVVKNTEFYKFNKIDIKLNKLNHNHYTKYFNDDRKKLFFYFNVKNLNVETKYSC

GILSYENVNYYKKOKNVKFSYLFGLSPOSEGPKGKCLYKNAPLYKNYENKQYCNPPH  
YYPVYTFLESDDKIINEKNQERNKERNERNEKNKNGKNEKNDLSTTNNNVVMK  
NGIQSTHNNNSINSNDQKKNNNNNNYYIYNNPGLITNVKTKIRGVDATALLNSVK  
YNNLINDNLKQNFINTFVDLDSFNIEYNEWYTFSSYLEYQFNSFTLNNADLLK  
KNNLTPRYDEINELNKEIREITMEYKTDKEYQYOVNNEYINENGVDEKGYKTKELK  
LNDFTYIKPMYKMSKINAFDETFKNIRFVFLVCLGVFINVYKIDINERKKTINENFL  
CCLKINKYVYVYWSLLFYILFFVYINFTLVVYVYVYKLANFYTLFLVYIFVFLANS  
LDTITICHOFSNNSINIYATLFLFPLFSFRLLIHSGSANGITLTPVLLIHPHSSFCPLS  
LDFILILKNNTIDYKOLFIFKENILSMHLIICSTSYFVLVILICILYIYKRRKKN  
MLNTYRGKSKITNDKRRKKSQKTYEYCNINYDMSEKYPTHQLKDSNDPDMLEGSDE  
DDTCYELDTLHKQVNSQOKLEKIRINDRESNGHPYDIPMLPHNNNNNNNNINNN  
INNN  
NGEYKSTLIITIKMSKDSGEINFKNYSKKNKDDIYVANKNVRNGLCGRICSNY  
IKFEYNNIFKRECNNTDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
KNDLSEYCSQVILYENLTFFYIKFILLYYKNVYDKYLNKRTKLIMNDLQVYL  
NDQIKNSDQVKKKITSICFLVFKRDIYILDEPFTALDITKTKFLPKFDPKTKNNI  
FICTHDIYEAANFANDIAVKSQIIFKGSKNFSOKLIEYKFTLNQIQPNNEEOTN  
YLSQBKISANLKNKGVIKSGQVANNYVDGVNVRISSESSINOKKNDNNNDND  
NNKNSVDIHDLPDILVNTINFTNNREDIIEKNILIKFTKPTNENKCFIENENHI  
YCTYKILNELESKLLKLLVCNKKFNILTYQLKTDIYYIYIYTLNEKKKLLKNIQDK  
GLDILDEPFLFFQNFQPNFLNLLKNNNHMQPIISYFAYLRDLVLVGNENKN  
LIINSHDNDPYDIJNNNNHKKYIKNNNNNSYNNRDDTELDEIENISSKKKN  
KIEKFYSYTYNNKNNDEGEYDNNNDMMNSNTSNMKKKTKKKNQIKFTCI  
NFTTYIKPTLLKLDKGLSGSYWYKFTVPLFLFGLGLIICVFSYGVQNDLQ  
YSTYISGNLKQSTINVTLYKNEMANNFTLYKNYLNDDNTIKNLNSYDKLKKLE  
TKQDQNAEESHINNNDNDIKDLKENNEYNKLHENSEFNYSNTINSLOKCY  
LKEKNTYDEEILDNDVDSINLYLEKANHENEJSLGTFAPLITPEKADSNDEY  
ETLFTNDYDNKVEIINILNYPYTSIHSYAYTNSLPMLSQDNTLNKSKSGKNI  
LDGTSYNDKVVEDYKQSDNNTLIDKENKYNILIDKDNKKEBENPLVKKLRKEEDNAFRUKN  
VSYNNNEEDINGNINDSNNIYITPKNNKKEBENPLVKKLRKEEDNAFRUKN  
FHEPMRPFYINIVYFLSIVFCEFFCFMFEFKNEIKRNFENHFHQYIHYFOILLLE  
YLYPFIYTLCLFVILFYIDFKBFPSFFCFLLYGLFNILFICLSYLSHSYILFL  
FNFNFKGTHIDIVLTVLSYVANNHEILINLSHVLCIFRIFDSFALSHNNIRSLC  
LNVKRMKHIDIEDIYEDWMSNNYVEIFGCFKKVYNQSGDNNLISDGIQSVCEDSN  
SFNTSGDFVFLINCIYIYLSYVFKLQIIPSKERTDKKNESEKTYEELIILNYSN  
VGQBRKEQKNRYTFLLKNYFLSKKEYKQKPEYKKNPFPFYLOKLINPGRGNY  
NDNTMTGVIYKGEKVDVSNYSDEEININQOLNSEGYLKNDIYNELNDNYNDPDD  
NTVDIEMQISEKSNIKNLYKETIDKYIKGSDGSDNDDEKNFNSOKYILKNLNI  
KMRPYITFSSYNNNDLNIYFFYFICFNNKNDKLSQTIAYGQYIEINKFNKFNKK  
NONEHISEKIOKMRDEYKIELIEMPTSYHEPKIILTFKNIETNVEIKYIYNILMLI  
NWKCDIHINCSGNLGMKKVELMINLRDDKIIFLYKLDNDIDFCSQYIYNILMLI  
ILLYNCEDEHKGDEYKDVYKQYKDEYKDKYKDEYKDYKOTYKDEYKVEYED  
RVYEQDIOLYKGEYITNDNOFDYIDKYSILYDLRGDEKDEKOLNEMLRGHRKNY

	Query Match	3.4%;	Score 47.8;	DB 3;	Length 253305;
	Best Local Similarity	61.8%;	Pred. No. 0.25;		
	Matches	76;	Conservative	0;	Mismatches 47; Indels 0; Gaps 0;
QY	1266	tgttattgagcacttcgtggacaagccaagtatctttataaaattttatagcatttagta	1325		
Db	78852	TTTTATTCTTCATTTTGGTTAAAAAGATGATATTTTATAAATATTATATTTATGTAATA	78911		
QY	1326	tacttggtatatgtgttttttattgaagctgtggaa	1385		
Db	78912	TAATTAATCTATGTTTTTTTTTTTTTTTTTTTTTTTAAAGTATGAATATATAATATAATATA	78971		
QY	1386	aaa	1388		
Db	78972	ATA	78974		

RESULT	12	
AF151733		
LOCUS		
DEFINITION		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS		

Jermyn,K.A., Mohanty,S., Firtel,R.A. and Williams,J.G.  
 An F-Box/WD40 repeat-containing protein important for Dictyostelium  
 cell-type proportioning, slug behaviour, and culmination  
 Dev. Biol. 224 (1), 42-59 (2000)  
 20359235  
 10898960  
 2 (bases 1 to 3885)  
 Nelson,M.K., Clark,A., Abe,T., Nomura,A., Firtel,R.A. and  
 Williams,J.G.  
 Direct Submission  
 Submitted (14-MAY-1999) Biology, Allegheny College, 520 North Main  
 Street, Meadville, PA 16335, USA  
 Location/Qualifiers  
 1. 3885  
   /organism="Dictyostelium discoideum"  
   /db\_xref="taxon:44689"  
   /chromosome="2"  
   /map="position 6110"  
   /join("<1..237,379..>3885")  
   /gene="fbxA"  
   /product="F-Box A protein"  
   <1..>3885  
   /gene="fbxA"  
   /join(1..237,379..3885)  
   /gene="fbxA"  
   /function="regulate cell fate"  
   /note="FbxA"  
   /codon\_start=1  
   /product="F-Box A protein"  
   /protein\_id="AAD40673.1"  
   /db\_xref="GI:5163418"  
   /translation="MQYVNGNDISQSITHEVLEAFSTKIPTWKLEYESSKHHIHSY  
 SSHERGYLSKGVAEFNPSNPVYELFLDIISSRSKWDHCFHREAHIEEYDHLNHIIL  
 NFTPFLISNLDMNLYRSCKYDPQERLEFVIAMRSIELEDGDVDQFELPNGVIGLGR  
 QDKCKMTFVQOCDLTELQIRPGYKFSKLERLEDQFLTLFPATVSGRLAKIFES  
 TELXISNMVGIETKDRIPTISIMEKAEKVENMEFGTNPDYGMKIVLKLDEILIKTK  
 TSGYIMIGKGSFSSRYPELLADVLVLYOKNPFEMDTFYDKTLKLVESINREVEVHYR  
 MWRNTIWRLLQSVKKGSGFSSVHHRSICSPNQVADGIEVHILPTALLYGLGDGS  
 FTSFLAALEVGQYPTPWEEMWMTFAARIISNQNSIINHVNKLTGSGEMIRELPVPI  
 VQHHGDGVEGSHITGPVYNAMVENADILLDEGNKDGVSXSKRNRKNRKNENKNN  
 SVDNIIDEEIEEENIEKDQOQTQQOQTQQOQQOQQOQTQQOQQOQQOQQOQQOQQ  
 NSHOLKKEIKPTPSIRYLLLSNSGSRPIAWYESKKOPFLFVSDSSRRKGIR  
 KTYVFSNSGPNLDEEVQIILISNLSAINIVNLVLCRKREKMATDSPILWKNLYKSNP  
 LFHKKTPARKQIILISLNDENKYNNSGDSADGSSSEEEQQOQQOQQOQQOQQOQQ  
 QOQQTFDSLNLNDLPDIAKTOHLSMISQNPATIQCESPINKNSGNDPKPNKRNDR  
 TQOLPHQVLSLIQWNTDLAQHLSMISQNPATIQCESPINKNSGNDPKPNKRNDR  
 NSYIPDEEFINWKGSHYTBKHKQSRKRWMEIRITPLKGNRAIKAVKSEGNSTAVT  
 TEKKIKFWNLNTDEQFIEGESGVLSYEDTKQSCIMPLSDYTKVHCHKNGTGT  
 MVDTEQPIEVHTSRPNLADGDFTEPGKYLWEHTIIHYWDETSTLLWNLNLAH  
 TRKTTQSKVAQHLSLKNKGIYFTVSSDKSAKWDLTNGCTINLVGHISYAVNCIEPIG  
 DYMALGTQKATLNDQURATFISSTYSGTGPTRICISYOEKNGIVLSDGDSGSI  
 AFNLDNNLNSIVNIGVKPIFNKNLIGTTTTTTTTTTTTTTTTTTTTTTNTNTNVP  
 INLGTGDKSLRHHNESAVTCIESDEAGFISGQNLGLRWDF"  
 1507 a 567 c 602 g 1209 t

Query Match	3.2%;	Score 44.6;	DB 3;	Length 3885;
Best Local Similarity	68.1%;	Pred. No. 0.95;		
Matches	62;	Conservative 0;	Mismatches 29;	Indels 0; Gaps 0;
QY 1298	tcttataaataatctatagcattagaggtatactgtgtatatctgtgtttttattaaagctgtg	1357		
Db 249	TATTATTATTATTAATTTTAAATTTTAAATTTTATTTTATTTTATTTTATTTATTT	308		
QY 1358	gaataaataatattataaaaaaiaaaaaa 1388			
Db 309	TATTTAAATAAATAAAGAAAAAIAAAAAA 339			
RESULT 13				
AF151111	3985 bp	DNA	INV	29-MAR-2000
LOCUS				

WE CHINA LINE CONG/WASING/ABI CONSULTING FOR SEQUENCING BAC CLONES  
F6P23, F5J6, and T13L16, the ESSA group for sequencing clone  
F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards

and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

```

FEATURES
    source
        Location/Qualifiers
            1..110211
                /organism="Arabidopsis thaliana"
                /cultivar="Columbia"
                /db_xref="taxon:3702"
                /chromosome="II"
                complement(<1..2793)
                /note="Sequence from clone F27L4"
                complement(356..401)
                /rpt_family="(TAA)n"
                complement(join(1549..2058,2731..3012))
                /gene="At2g23810"
                complement(1549..3012)
                /gene="At2g23810"
                /note="F27L4.1"
                complement(join(1744..2058,2731..3003))
                /gene="At2g23810"
                /codon_start=1
                /product="putative senescence-associated protein 5"
                /protein_id="AAFI8611.1"
                /db_xref="GI:6598449"
                /translation="MFLILLVFCITVFAFVTVNKGAGEAIEGKYKEYKLGDYSTWL
                QKRVNKNKIRKCLVESKVCSEKLEAKFVNVPSFYKEHLTALQSGCCPKPSDECG
                FEVNVPTWTKNTGTHTNPDCOTWDNAKEKLCFDCQSCRAGLLDNVKSARKKVAIYN
                IYFLVLLIIVSYGCCAFRNKRDDSYRYYGYP"
                /rpt_family="(CAA)n"
                /note="Sequence from clone T29E15"
                complement(3567..3677)
                /rpt_family="(TAAA)n"
                complement(3570..3641)
                /rpt_family="(TAAA)n"
                complement(4565..4612)
                /rpt_family="(CAA)n"
                /note="T29E15.2"
                /rpt_family="(CAAAA)n"
                /join(<6480..6749,7602..7726,7811..7848,7929..8039,
                8152..8268,8487..8557)
                /gene="At2g23820"
                <6480..8786
                /gene="At2g23820"
                /note="T29E15.2"
                /join(6480..6749,7602..7726,7811..7848,7929..8039,
                8152..8268,8487..8557)
                /gene="At2g23820"
                /note="unknown protein"
                /codon_start=1
                /protein_id="AAC63656.1"
                /db_xref="GI:3738315"
                /translation="MNKNRVFFCKLSLVFLVPSFTRSHIRFTYSNAGASSNRAIH
                CMASDPQSGDGVSPPNVAAPSSSSSSASDAIDFLSLCTRLKTTPRAGWTKRDV
                KDPESTADHRLMALISDIPGVNRDKCMKMAIVHDAEAIIVGDITPSCGISKEE
                KNRESEALHMKLLGGGRAKEIAELWREYBENSPEAKVVKDFDKVELILQALEY
                EOCKFOTNICKAWASEYRRRKQH"
                complement(6654..6712)
                /rpt_family="(GAA)n"
                /join(<9531..9582,9709..9779,9866..9947,10054..10169,
                10268..10396)
                /gene="At2g23830"
                <9531..10396
                /gene="At2g23830"
                /note="T29E15.3"
                /join(9531..9582,9709..9779,9866..9947,10054..10169,
                10268..10396)

misc_feature
    repeat_region
    repeat_region
    repeat_region
    repeat_region
    repeat_region
    mRNA
    gene
    CDS

repeat_region
    repeat_region
    repeat_region
    repeat_region
    repeat_region
    mRNA
    gene
    CDS

```

```

/gene="At2g23830"
/note="unknown protein"
/codon_start=1
/protein_id="AAC63657.1"
/db_xref="GI:3738316"
/translation="MSNNELLEIEPMYLQFFELKQKSCSLYLTNKTNDNNVAFKVKT
TNRNYSVRPNYGLILPKSTCKVLVTOAQRPVSDMQSFKEFKTIQSYLASPGVTAKE
VTRMFSGSHVVEETKLRVYVCSTTTNTSSPTRRGEIFQCFC"
complement(11332..11371)
/rpt_family="(TAAA)n"
15801..15917
/rpt_family="(TAAA)n"
15853..15918
/rpt_family="(CAAAA)n"
16051..16136
/rpt_family="(TAAA)n"
complement(join(<20442..20537,20617..20691,20894..20948,
21145..21485))
/gene="At2g23840"
complement(<20442..21485)
/gene="At2g23840"
/note="T29E15.4"
complement(join(20442..20537,20617..20691,20894..20948,
21145..21485))
/gene="At2g23840"
/note="unknown protein"
/codon_start=1
/protein_id="AAC63658.1"
/db_xref="GI:3738317"
/translation="MAHFSARKLLSSNDGLSGFVDSRQFRQSLYMAGGSPVPIR
MSRLKFCARIFSSYFPYKRKKIRDLDNTEKNLDIDEDNEDWDFDGGDDGLTDHLS
CFRLGLDISYRPNVVCWKRAICLEYMDKADVLEYVDQTVSSPTGSGYIPAVLRVPH
LLOVKKRRVKNLSRKNIILLRDDYTCQ"
complement(22086..22781)
/gene="At2g23850"
/note="T29E15.5"
/pseudo
join(<22999..23070,23189..23296,23402..23554)
/gene="At2g23860"
<22999..23554
/gene="At2g23860"
/note="T29E15.6"
join(22999..23070,23189..23296,23402..23554)
/gene="At2g23860"
/note="unknown protein"
/codon_start=1
/protein_id="AAC63659.1"
/db_xref="GI:3738318"
/translation="MTNPNKYVRPNYGLILPKSTCKLKRKPGSPDMQSNKEKFIQIR
VKASPAVTAKEVLTLETFNKESGHLVEETKLRVYVCSTTTNITSPPTRNGLYFNVFY
RHFLFTSL"
27493..27571
/rpt_family="(TAAA)n"
complement(29076..29146)
/rpt_family="POLY_A"
29302..29885
/gene="At2g23870"
/note="T29E15.7"
/pseudo
31472..31519
/rpt_family="(TA)n"
complement(join(<31877..32176,32267..32407,32498..32893,
33162..35713,35886..36147))
/gene="At2g23880"
complement(<31877..36147)
/gene="At2g23880"
/note="T29E15.8"

```

Query Match 3.2%; Score 44.6; DB 8; Length 110211;  
 Best Local Similarity 63.6%; Pred. No. 1.5;  
 Matches 68; Conservative 0; Mismatches 39; Indels 0; Gaps 0;  
 Qy 1282 gtgacaagccaagtatctataataattatagcattagagtagtattgttatgttg 1341

DB 64715 GCTACAGTCAATGAATAAGATTTCAATTAATAAAAAAATCCATATTAGTTA 64774  
QY 1342 tttttttaagctgtggataataattattataaaaaa 1388  
DB 64775 TTTTGTGAGCGACATAATGAATATTATTATAAAAAATTACAACAA 64821

RESULT 15

SCOR17/c  
LOCUS SCOR17 366 bp DNA PLN 26-FEB-1992  
DEFINITION S.cerevisiae DNA for ori 7.  
ACCESSION X59536  
VERSION X59536.1 GI:14234  
KEYWORDS ORI region.  
SOURCE baker's yeast.  
ORGANISM Mitochondrion Saccharomyces cerevisiae  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
REFERENCE 1 (bases 1 to 366)  
AUTHORS Skelly,P.J. and Clark-Walker,G.D.  
TITLE Sequence rearrangements at the ori 7 region of Saccharomyces  
cerevisiae mitochondrial DNA  
JOURNAL J. Mol. Evol. 32 (5), 439-442 (1991)  
MEDLINE 91251143  
FEATURES  
source Location/Qualifiers  
1..366  
/organism="Saccharomyces cerevisiae"  
/organelle="mitochondrion"  
/db\_xref="taxon:4932"  
BASE COUNT 188 a 25 c 28 g 125 t  
ORIGIN

Query Match 3.2%; Score 43.8; DB 8; Length 366;  
Best Local Similarity 66.3%; Pred. No. 1.1;  
Matches 63; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
QY 1294 agtatcttataaattattatagcattagtagtactgtgtattgtttttttaagc 1353  
DB 241 ATTATATCTTAAATTTTATTATATATATATATATATATATATATATATATAT 182  
QY 1354 tgtggaataaataattattataaaaaa 1388  
DB 181 TATATAATATTATATATTTTATTATAAAAAATTATA 147

Search completed: February 28, 2002, 00:56:47  
Job time: 5649 sec

**THIS PAGE BLANK (US--)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2002, 23:59:59 ; Search time 113.25 Seconds  
(without alignments)  
10507.439 Million cell updates/sec

Title: US-08-816-011f-36  
Perfect score: 1388  
Sequence: 1 atgtgtaataatcccgatc.....ttattaaaaaaaaaaaaa 1388

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 930621 seqs, 428662619 residues  
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%

Database : N\_Geneseq\_1101.\*  
Listing first 45 summaries

- 1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1386.4	99.9	1388	17 AAT18168	CORK potassium cha
2	85	6.1	936	22 AAF58252	Oligonucleotide D1
3	85	6.1	936	22 AAF58254	Oligonucleotide D1
4	85	6.1	936	22 AAF58257	Oligonucleotide D1
5	85	6.1	936	22 AAF58259	Oligonucleotide D1
6	85	6.1	936	22 AAF58262	Oligonucleotide D2
7	85	6.1	938	22 AAF58255	Oligonucleotide D1
8	83	6.0	936	22 AAF58252	Oligonucleotide D1
9	83	6.0	936	22 AAF58254	Oligonucleotide D1
10	83	6.0	936	22 AAF58257	Oligonucleotide D1
11	83	6.0	936	22 AAF58259	Oligonucleotide D1

C 12	83	6.0	936	22 AAF58262	Oligonucleotide D2
C 13	83	6.0	938	22 AAF58255	Oligonucleotide D1
C 14	43.2	3.1	244	22 AAF58238	Oligonucleotide D1
C 15	42.4	3.1	244	22 AAF58238	Oligonucleotide D1
C 16	39.4	2.8	362	22 AAF5894	Novel human polynu
C 17	39.2	2.8	440	18 AAT67765	H. pylori secreted
C 18	39.2	2.8	440	18 AAT77445	H. pylori secreted
C 19	38.6	2.8	310	22 AAT71505	Human cervical can
C 20	38.6	2.8	1558	17 AAT28255	Survival motor neu
C 21	38.6	2.8	1560	17 AAT18828	Human survival mot
C 22	38.6	2.8	1582	17 AAT28259	Survival motor neu
C 23	38.6	2.8	1582	17 AAT18831	Human survival mot
C 24	38.4	2.8	2439	20 AAX80667	Clone am856_3 enco
C 25	38.4	2.8	10732	21 AAA10594	Gene encoding a su
C 26	38	2.7	3499	21 AA255832	Staphylococcus aur
C 27	38	2.7	4709	18 AAV74898	Staphylococcus aur
C 28	37	2.7	2079	21 AAA26360	Human secreted pro
C 29	36.8	2.7	65792	22 AAF28544	Genomic fragment #
C 30	36.6	2.6	513445	22 AAI61373	Soybean 318013 reg
C 31	36.4	2.6	1338	21 AAC93411	Human secreted pro
C 32	36.2	2.6	1667	22 AAF85681	Pea blight resista
C 33	36.2	2.6	1851	21 AAC79950	Human secreted pro
C 34	36.2	2.6	3076	22 AAH46951	Human secreted pro
C 35	36.2	2.6	5761	22 AAI59388	Human polynucleoti
C 36	36	2.6	1077	19 AAV39085	Human tumour necro
C 37	36	2.6	1077	21 AAA37772	Human tumour necro
C 38	36	2.6	1164	19 AAV07654	Nucleotide sequenc
C 39	36	2.6	1179	22 AAH33171	Human colon cancer
C 40	36	2.6	2215	15 AAQ70882	Tomato spotted wil
C 41	36	2.6	2621	16 AAQ76109	Tomato spotted wil
C 42	36	2.6	2916	12 AAQ11655	TSWV S RNA. Tomat
C 43	36	2.6	2916	22 AAC89646	Tomato spotted wil
C 44	36	2.6	2964	22 AAF88022	Cucumber LBLOX DNA
C 45	36	2.6	2970	20 AAX86461	RNA encoding non-s

ALIGNMENTS

RESULT 1  
AAT18168  
ID AAT18168 standard; cDNA: 1388 BP.

XX AAT18168;

XX 16-OCT-1996 (first entry)

XX CORK potassium channel gene.

XX CORK; potassium channel; nematode; polyadenylation site;

KW potassium dependence; Saccharomyces cerevisiae; potassium-agonist;

KW potassium-antagonist; drug screening; nematocide; anthelmintic;

KW cardiac disorder; ss.

XX Caenorhabditis elegans.

XX Key Location/Qualifiers

FT CDS 1..1305

FT /tag= a

FT /product= CORK potassium channel

FT polyA\_signal 1359..1364

FT /tag= b

XX WO9613520-A1.

XX 09-MAY-1996.

XX 25-OCT-1995; 95WO-US14364.

XX 31-OCT-1994; 94US-0332312.

XX (AMCY ) AMERICAN CYANAMID CO.





















```
Qy 702 ggaattcgcaaaaggcggaacagacagacgacctctccattctttggaccacatt 761
Db 717 WWWWWWNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Qy 762 cacaactgttatggcgcaactcttcaatgtttggtctgcttgcggtactctcacaat 821
Db 657 WWWWWWNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Qy 822 ctccctgttatgatgacggttaccaactcgtggagatccggtctcctaaacaaattat 881
Db 537 WWWWWWNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Qy 882 gtctgaacacgatgaattacacatttctcacaagtttctcgtcttcaatttggctgc 941
Db 537 WWWWWWNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Qy 942 tgcgattgaccatagttgttccaaagattcactgcgcgacaccccttaccctcaatt 1001
Db 477 WWWWWWNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Qy 1002 tgcacaaattcgtcgtctcttccattccattcttctcgtcgaactatcgtgtcca 1061
Db 417 WWWWWWNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Qy 1062 gacgcgtcttatcctgttttcttggtagtactgacatttttggattggtggaattgc 1121
Db 357 WWWWWWNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Qy 1122 catgtcttttcaactgatacctcagcgtctcggcaatgggatacacactcacaacgtcgt 1181
Db 297 WWWWWWNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Qy 1182 gcaactcactactcaagatttgcgcgtcagcttccgttttgcactcttatggttggcct 1241
Db 237 WWWWWWNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Qy 1242 tctcaccggtgacctgtgcccgttattgagcactcgtggcacaagccaagtattctt 1301
Db 177 WWWWWWNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Qy 1302 ataaattattatagcattagatagatacattgttatgtttttattaaagctggaat 1361
Db 117 WWWWWWNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Qy 1362 aaataattattaaaaaiaaaaaa 1388
Db 57 WWWWWWNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

```
RESULT 14
AAF58238/c
ID AAF58238 standard; DNA; 244 BP.
```

```
XX AAF58238;
XX
XX
XX 24-APR-2001 (first entry)
XX
XX Oligonucleotide D1250:D1102.
XX
XX Electron-transfer group; ETM; mismatch; genotyping;
XX gene expression; ss.
```

```
XX Synthetic.
```

```
XX WO200107665-A2.
```

```
XX 01-FEB-2001.
```

```
XX 26-JUL-2000; 2000WO-US20476.
```

```
XX 26-JUL-1999; 99US-0145695.
```

```
XX 17-MAR-2000; 2000US-0190259.
```

```
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI Umek RM;
XX
XX WPI; 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
XX
XX Example 4; Page 120; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX
XX Sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other;
```

```
Query Match 3.1%; Score 43.2; DB 22; Length 244;
Best Local Similarity 8.0%; Pred. No. 0.057;
Matches 12; Conservative 95; Mismatches 43; Indels 0; Gaps 0;
```

```
Qy 1239 cctctcaccggtgacctgtgcccgttattgagcactcgtggcacaagccaagtat 1298
Db 226 CCTCATTGATGCTGCTTTTAAACAWNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Qy 1299 cttataaatattatagcattagatatactgttatgtttattattaaagctgtgg 1358
Db 166 WWWWWWNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Qy 1359 aataaaataattataaaaaaiaaaaaa 1388
Db 106 WGWNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

```
RESULT 15
```

```
AAF58238
ID AAF58238 standard; DNA; 244 BP.
```

```
XX AAF58238;
```

```
XX
XX
XX 24-APR-2001 (first entry)
```

```
XX
XX Oligonucleotide D1250:D1102.
```

```
XX
XX Electron-transfer group; ETM; mismatch; genotyping;
XX gene expression; ss.
```

```
XX Synthetic.
```

```
XX WO200107665-A2.
```

```
XX 01-FEB-2001.
```

```
XX 26-JUL-2000; 2000WO-US20476.
```

```
XX 26-JUL-1999; 99US-0145695.
```

```
XX 17-MAR-2000; 2000US-0190259.
```

```
XX (CLIN-) CLINICAL MICRO SENSORS INC.
```

```
XX Umek RM;
```

```
XX WPI; 2001-159728/16.
```

```
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
```

[illegible]

Search completed: February 28, 2002, 00:52:31  
Job time: 3152 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2002, 23:21:59 ; Search time 1212.08 Seconds  
(without alignments)  
12305.409 Million cell updates/sec

Title: US-08-816-011f-36  
Perfect score: 1388  
Sequence: 1 atggtaataatcagcagatc.....ttattataaaaaaaaaaaaaa 1388

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Existing first 45 summaries

Database :

EST:\*\*  
1: em\_estfun:\*\*  
2: em\_esthum:\*\*  
3: em\_estin:\*\*  
4: em\_estom:\*\*  
5: em\_estpl:\*\*  
6: em\_estba:\*\*  
7: em\_estro:\*\*  
8: em\_estov:\*\*  
9: em\_htc:\*\*  
10: gb\_est1:\*\*  
11: gb\_est2:\*\*  
12: gb\_htc:\*\*  
13: gb\_gss:\*\*  
14: em\_gss\_fun:\*\*  
15: em\_gss\_hum:\*\*  
16: em\_gss\_inv:\*\*  
17: em\_gss\_pln:\*\*  
18: em\_gss\_pro:\*\*  
19: em\_gss\_rod:\*\*  
20: em\_gss\_vrt:\*\*  
21: em\_gss\_other:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	86.6	6.2	533	10	AU222343
c 2	77.6	5.6	332	10	AU201199
c 3	75.6	5.4	710	10	AU216992
4	75.2	5.4	648	10	AW453425
5	73.8	5.3	327	10	AU109268
6	73.8	5.3	642	10	AI082929
7	70.8	5.1	495	10	AU209436
8	70.2	5.1	360	11	CI3829
c 9	67.8	4.9	300	10	AU113471
c 10	57.6	4.1	300	10	AU115402
c 11	57.4	4.1	300	10	AU114995
12	56.2	4.0	997	13	CNS0057E

13	54.4	3.9	258	10	AA406898
14	53	3.8	656	10	AW409482
15	50.4	3.6	345	10	AU109592
16	49	3.5	606	10	AA948920
c 17	48.8	3.5	901	13	CNS0760F
18	48.2	3.5	1201	13	CNS0010J
19	47.6	3.4	274	10	AA842090
20	47	3.4	228	10	AI784877
21	46.8	3.4	729	10	AI239052
22	46.2	3.3	360	11	R03450
23	46.2	3.3	373	11	C43216
24	45.8	3.3	274	10	BB181657
25	45.6	3.3	350	10	AU111089
26	45.6	3.3	351	10	AU111161
27	45.6	3.3	738	13	AQ842781
c 28	45.6	3.3	759	13	AQ254298
29	45	3.2	375	11	C44886
30	44.6	3.2	340	10	AU111560
31	44.6	3.2	340	10	AU111734
c 32	44.6	3.2	510	13	AQ254261
c 33	44.4	3.2	429	10	AW829621
c 34	44.4	3.2	448	10	AW829541
c 35	44.2	3.2	333	10	AI431074
c 36	44.2	3.2	441	11	BF459052
37	44.2	3.2	459	10	AI595433
38	44.2	3.2	475	10	AA058117
c 39	44.2	3.2	566	10	AW333515
c 40	44.2	3.2	626	10	AW048063
41	44.2	3.2	1101	13	CNS012EM
42	44.2	3.2	2664	12	AK005136
c 43	43.6	3.1	227	11	BF460904
44	43.6	3.1	234	10	BB169820
45	43.6	3.1	327	10	BB182391

#### ALIGNMENTS

RESULT 1  
AU222343/c

LOCUS AU222343 533 bp mRNA EST 17-JUL-2001  
DEFINITION AU222343 unpublished oligo-capped cDNA library, stage L1  
Caenorhabditis elegans cDNA clone yk1015b04 3', mRNA sequence.

ACCESSION AU222343

VERSION AU222343.1 GI:14860500

KEYWORDS EST

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans

REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea

AUTHORS ; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 533)

Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.

A Sugano,S.

A complementary view of the C.elegans genome

Unpublished (2001)

Contact: Yuji Kohara

Genome Biology Lab.

National Institute of Genetics

Yata lili, Mishima, Shizuoka 411, Japan

Tel: 81-559-81-6854

Fax: 81-559-81-6855

Email: ykohara@lab.nig.ac.jp.

Location/Qualifiers

1..533

/organism="Caenorhabditis elegans"

/strain="N2"

/db\_xref="taxon:6239"

/clones="yk1015b04"

/clone\_lib="unpublished oligo-capped cDNA library, stage

Li"

/sex="Hermaphrodite"

/tissue\_type="whole animal"

/dev\_stage="L1"

FEATURES

source

```
BASE COUNT      191 a   101 c   100 g   141 t
ORIGIN

Query Match      6.2%; Score 86.6; DB 10; Length 533;
Best Local Similarity 49.3%; Pred. No. 7.9e-08;
Matches 255; Conservative 0; Mismatches 259; Indels 3; Gaps 1;

QY 808 gttactctcaacaatcttcctgttatgagccgttaccactcgtggagattccggcttc 867
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 GTTCTCTCTCTCATCTTCTCTGCAATGATGACATGATAGTGTACTGTGATCTACAAAT 455

QY 868 ctaaacaaaattatgtctgaacacgatgaatctacactttgtctcacaagttctcctcgtc 927
    || || || || || || || || || || || || || || || || || || || || ||
Db 454 GGAATAACTTCTCTTTTGGAGATAGTTTATCTCTCC---GATAACTACTTTCCCTCAAT 398

QY 928 ttcaatttgcctgcgattgagatccatagttgttccaaagattccactggccgacacc 987
    ||||| ||| || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 397 TTCAATTTATTTGCATGGATTGGATCATCTCTTGGCAATTTATTTCAATTCOCATCAGAG 338

QY 988 cgttactcaaaatttgcataactctgcgtctcttcttccattccattctctctctcgc 1047
    || || || || || || || || || || || || || || || || || || || || ||
Db 337 AATATTTATGATGATGAGTGTCTCTCGAACCGTATTCATTCCTCAATTTATCTTTCTGT 278

QY 1048 aactcgtgtccagacgogtgcattcctgttcttcttcttcttcttcttcttcttctg 1107
    || || || || || || || || || || || || || || || || || || || || ||
Db 277 AATTATCGTCAAAATCTAGAACATGGCCTGTTGGTTTAAAGAAATGAATGGTGGTCACT 218

QY 1108 attggtggaattgccatgtctcttcttccatgatacctcagcgtctcgtgcaatgggatac 1167
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 217 ATTGGATGTACCAATAATGGCATTTTACTTGTGGTTATGAGCAGATTTAGCGCTGATTTAT 158

QY 1168 actccaaagctgtgcacatctcactcactcaagatttgcgcctcagcttcttccgttgcact 1227
    || || || || || || || || || || || || || || || || || || || || ||
Db 157 ACACCAAGCAAAAGTCCGCTAGATATACAACTAAGTGAATGCTTGCATCAATTTTC 98

QY 1228 cttatggttgcctctcaccggcctgtgcccgttgttattgagcaactcgtcgagac 1287
    || || || || || || || || || || || || || || || || || || || || ||
Db 97 CTAATCTCGGAATTTCTATCGGAGTCGCCAGTACACCCATTGCTGCATGGCGGTGGAC 38

QY 1288 aagccaagtatctataaatattattagcattagagt 1324
    || || || || || || || || || || || || || || || || || || || || ||
Db 37 TCATAGGAGTAGAAGAACATGACTGTATATAATT 1

RESULT 2
LOCUS      AU201199      332 bp      mRNA      EST      17-JUL-2001
DEFINITION AU201199 unpublished oligo-capped cDNA library, stage L4
            Caenorhabditis elegans cDNA clone yk789e10 5', mRNA sequence.
ACCESSION  AU201199
VERSION     AU201199.1 GI:14829871
KEYWORDS    EST
SOURCE      Caenorhabditis elegans.
            Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 332)
AUTHORS     Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
            and Sugano,S.
            A complementary view of the C.elegans genome
            Unpublished (2001)
            Contact: Yuji Kohara
            Genome Biology Lab.
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
            Location/Qualifiers
                1. 332
                /organism="Caenorhabditis elegans"
                /strain="N2"

BASE COUNT      191 a   101 c   100 g   141 t
ORIGIN

Query Match      5.4%; Score 75.6; DB 10; Length 710;
Best Local Similarity 51.6%; Pred. No. 1.2e-05;
Matches 199; Conservative 0; Mismatches 184; Indels 3; Gaps 1;

QY 949 ggaatccatagttggttccaaagattccactggccgacaccccggttaccctcaaatggcata 1008
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    /db_xref="taxon:6239"
    /clone_lib="unpublished oligo-capped cDNA library, stage
    L4"
    /sex="Hermaphrodite"
    /tissue_type="whole animal"
    /dev_stage="L4"

BASE COUNT      100 a   59 c   81 g   91 t   1 others
ORIGIN

Query Match      5.6%; Score 77.6; DB 10; Length 332;
Best Local Similarity 65.3%; Pred. No. 6e-06;
Matches 113; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 7 ataataccagatcacacacattatcgcttgagcagaaagcatttccaaagacaaagtac 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114 ATAGTTGACGGTGGGATAGCTCTGATGACACAAATGAGCTGATTCAGAAAGATAAGGGA 173

QY 67 aatattgtctactggctcgtcattcttggattcggagcttcttgcgatggaatag 126
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 CGTATGTTTTTGGATTATTTGCTCAATGGAATGGAGTTCTTTTGGCATGGAATAG 233

QY 127 ttoattactatgcgcctcgtgatttatttgaaattatttggttcaaacccgagtg 179
    || || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 234 TTATCATTATCGNCACAGTACTACGTGGATTACTGGTTTCACAGTAAATGG 286

FEATURES
            source
                1. 710
                /organism="Caenorhabditis elegans"
                /strain="N2"
                /db_xref="taxon:6239"
                /clone="yk844e08"
                /clone_lib="unpublished oligo-capped cDNA library, stage
                L4"
                /sex="Hermaphrodite"
                /tissue_type="whole animal"
                /dev_stage="L4"

BASE COUNT      251 a   103 c   181 g   172 t   3 others
ORIGIN

Query Match      5.4%; Score 75.6; DB 10; Length 710;
Best Local Similarity 51.6%; Pred. No. 1.2e-05;
Matches 199; Conservative 0; Mismatches 184; Indels 3; Gaps 1;

QY 949 ggaatccatagttggttccaaagattccactggccgacaccccggttaccctcaaatggcata 1008
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    /db_xref="taxon:6239"
    /clone_lib="unpublished oligo-capped cDNA library, stage
    L4"
    /sex="Hermaphrodite"
    /tissue_type="whole animal"
    /dev_stage="L4"

BASE COUNT      251 a   103 c   181 g   172 t   3 others
ORIGIN
```

and DNA pol I. The library was constructed in the lambda Uni-zap xr vector and has 1 x 10E6 independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigman@nybc.org)."

168 a 126 c 120 g 234 t

BASE COUNT  
ORIGIN

	Query Match	5.4%	Score 75.2	DB 10	Length 648
	Best Local Similarity	48.9%	Pred. No. 1.5e-05		
	Matches 305	Conservative 0	Mismatches 298	Indels 21	Gaps 3
Qy	30	tgcggttgacagagaagcattccaaagagacaagtacaatatgtctacttgccgtcgcat	89		
Db	28	TGATATGGTGTTCACAAACCTCCCAAGACAAGTACATGCTGTTTATTTCATTTTACT	87		
Qy	90	tcttgttggttcggaggtctctctgccatggaatatgtcttacttactatcgccctgagta	149		
Db	88	ACTTCACGGGATAGCGGCTGTGATGCCGTSGAACATGTTTCATCACAATGCTCCCATCTTA	147		
Qy	150	ttatgtgaattattggttc---aaaccggatggcggtggagacatcgg-----ta	194		
Db	148	CTACGTCAACTATATAATTCGTGGAAATGAATGCCGATGTTACGGTGCATAAAAGTGATTA	207		

	Query Match	5.4%	Score 75.2	DB 10	Length 648
	Best Local Similarity	48.9%	Pred. No. 1.5e-05		
	Matches 305	Conservative 0	Mismatches 298	Indels 21	Gaps 3
Qy	30	tgcggttgacagagaagcattccaaagagacaagtacaatatgtctacttgccgtcgcat	89		
Db	28	TGATATGGTGTTCACAAACCTCCCAAGACAAGTACATGCTGTTTATTTCATTTTACT	87		
Qy	90	tcttgttggttcggaggtctctctgccatggaatatgtcttacttactatcgccctgagta	149		
Db	88	ACTTCACGGGATAGCGGCTGTGATGCCGTSGAACATGTTTCATCACAATGCTCCCATCTTA	147		
Qy	150	ttatgtgaattattggttc---aaaccggatggcggtggagacatcgg-----ta	194		
Db	148	CTACGTCAACTATATAATTCGTGGAAATGAATGCCGATGTTACGGTGCATAAAAGTGATTA	207		

	Query Match	5.4%	Score 75.2	DB 10	Length 648
	Best Local Similarity	48.9%	Pred. No. 1.5e-05		
	Matches 305	Conservative 0	Mismatches 298	Indels 21	Gaps 3
Qy	30	tgcggttgacagagaagcattccaaagagacaagtacaatatgtctacttgccgtcgcat	89		
Db	28	TGATATGGTGTTCACAAACCTCCCAAGACAAGTACATGCTGTTTATTTCATTTTACT	87		
Qy	90	tcttgttggttcggaggtctctctgccatggaatatgtcttacttactatcgccctgagta	149		
Db	88	ACTTCACGGGATAGCGGCTGTGATGCCGTSGAACATGTTTCATCACAATGCTCCCATCTTA	147		
Qy	150	ttatgtgaattattggttc---aaaccggatggcggtggagacatcgg-----ta	194		
Db	148	CTACGTCAACTATATAATTCGTGGAAATGAATGCCGATGTTACGGTGCATAAAAGTGATTA	207		

	Query Match	5.4%	Score 75.2	DB 10	Length 648
	Best Local Similarity	48.9%	Pred. No. 1.5e-05		
	Matches 305	Conservative 0	Mismatches 298	Indels 21	Gaps 3
Qy	30	tgcggttgacagagaagcattccaaagagacaagtacaatatgtctacttgccgtcgcat	89		
Db	28	TGATATGGTGTTCACAAACCTCCCAAGACAAGTACATGCTGTTTATTTCATTTTACT	87		
Qy	90	tcttgttggttcggaggtctctctgccatggaatatgtcttacttactatcgccctgagta	149		
Db	88	ACTTCACGGGATAGCGGCTGTGATGCCGTSGAACATGTTTCATCACAATGCTCCCATCTTA	147		
Qy	150	ttatgtgaattattggttc---aaaccggatggcggtggagacatcgg-----ta	194		
Db	148	CTACGTCAACTATATAATTCGTGGAAATGAATGCCGATGTTACGGTGCATAAAAGTGATTA	207		

	Query Match	5.4%	Score 75.2	DB 10	Length 648
	Best Local Similarity	48.9%	Pred. No. 1.5e-05		
	Matches 305	Conservative 0	Mismatches 298	Indels 21	Gaps 3
Qy	30	tgcggttgacagagaagcattccaaagagacaagtacaatatgtctacttgccgtcgcat	89		
Db	28	TGATATGGTGTTCACAAACCTCCCAAGACAAGTACATGCTGTTTATTTCATTTTACT	87		
Qy	90	tcttgttggttcggaggtctctctgccatggaatatgtcttacttactatcgccctgagta	149		
Db	88	ACTTCACGGGATAGCGGCTGTGATGCCGTSGAACATGTTTCATCACAATGCTCCCATCTTA	147		
Qy	150	ttatgtgaattattggttc---aaaccggatggcggtggagacatcgg-----ta	194		
Db	148	CTACGTCAACTATATAATTCGTGGAAATGAATGCCGATGTTACGGTGCATAAAAGTGATTA	207		

AUTHORS	Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y. and Sugano,S.
TITLE	A complementary view of the C. elegans genome
JOURNAL	Unpublished (2000)
COMMENT	Contact: Yujii Kohara

/clone\_lib="Brugia malayi adult male cdNA (SAW94NL-BmAN)"  
/lab\_host="X11-Lambda MRF"  
/note="vector: Bluedo UniZap XR; Site\_1: EcoR I; Site\_2:  
Xho I; Lymphatic filarial nematode parasite of humans.  
mRNA was prepared from adult males of Brugia malayi  
isolated from jirds and converted to double stranded cDNA  
using reverse transcriptase and oligo(dT) followed by  
RNase H and DNAPol I. The library had 4.6 x 10E6  
independent recombinants and average insert size was 800  
base pairs. The library was constructed by Noelle Ling.  
The library is available from Dr. S.A. Williams, email  
genome@smith.edu"



JOURNAL COMMENT	Unpublished (2001) Contact: Yuji Kohara Genome Biology Lab. National Institute of Genetics Yata 1111, Mishima, Shizuoka 411, Japan Tel: 81-559-81-6854 Fax: 81-559-81-6855 Email: ykohara@lab.nig.ac.jp.														
	Location/Qualifiers														
FEATURES	source														
	1. 495 /organism="Caenorhabditis elegans" /strain="N2" /db_xref="taxon:6239" /clone="yk1015b04" /clone_lib="unpublished oligo-capped cDNA library, stage 1" /sex="Hermaphrodite" /tissue_type="whole animal" /dev_stage="L1"														
BASE COUNT		146 a 80 c 98 g 171 t													
ORIGIN															
Query Match		5.1%; Score 70.8; DB 10; Length 495;													
Best Local Similarity		51.0%; Pred. No. 0.00013;													
Matches 202; Conservative		0; Mismatches 182; Indels 12; Gaps 1;													
QY	39	gcagaagcatttcacagagacaaatgataattgtctactggtcgctgcttctgttg	98												
DB	110	GAAGAGAGTTCCACCAACAGATCGATGGTATCTTGTATATAATTTTACAAATGACAGG	169												
QY	99	attcgaggtctctgcagatgaatgttcaattactatcgccctgagattatgtgaa	158												
DB	170	AATGGGAATGTTAATGTCATGAATATGTTTATACAAATGCTCCACAATACTACTACGA	229												
QY	159	ttattggttcaaacggatggcgtggagacatggtattcgaagaattcatggatcttt	218												
DB	230	TTATTGGTTTAAACA-----ATACAAATTTATCAAGACAGTTTTATGTCAATTAT	277												
QY	219	gagcattggtctcaacacttccaaacgcaagcattgaatttttcaacctgttctcatat	278												
DB	278	TGGAGTGACGTACAGATCCAAATGTTGGAATATGATTTTAAATACAAATTTGGTCA	337												
QY	279	tgctggtccctgatctaccggtgtttgtctcgggtttgcttcaacatgcaacctgac	338												
DB	338	GCTAGGTTTCATGATGCTTCGAGTTGTGTTCTCTCATTTGGAATGCAATTTTGATTG	397												
QY	339	aatcattctcctctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	398												
DB	398	GGTCATGTTATCTTGGCATTTTGTAAACACCATCTCCAGACAGTTTACATGTTT	457												
QY	399	ctgggttaactcttggaatggcgacttcaatcaatt	434												
DB	458	TATTGTACTCTAATCATCATTAATGGCTATGAATTT	493												
RESULT 8															
LOCUS	C13829	360 bp		mRNA		EST		28-DEC-1998							
DEFINITION	C13829	Yuji Kohara unpublished cDNA Caenorhabditis elegans CDNA clone yk183c7 5', mRNA sequence.													
ACCESSION	C13829														
VERSION	C13829.1	GI:1561382													
KEYWORDS	EST.														
SOURCE	Caenorhabditis elegans.														
ORGANISM	Caenorhabditis elegans.														
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea														
AUTHORS	1 (bases 1 to 360)														
TITLE	Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano														
JOURNAL	M., Miyata, A. and Nishigaki, A.														
Expression map of the C.elegans genome															
Unpublished (1996)															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															
COMMENT															

```
/strain="N2"
/db_xref="taxon:6239"
/clone="yk708a4"
/clone_lib="unpublished oligo-capped cDNA library"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT      114 a 50 c 59 g 77 t
ORIGIN

Query Match      4.9%; Score 67.8; DB 10; Length 300;
Best Local Similarity 55.8%; Pred. No. 0.00061;
Matches 129; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Qy 1024 ttcttcattctctctctgaactatcggtgcacagcgctgttaccctgttttc 1083
Db 299 TTCATTCATTCTTCATGTTGAGCACTATTTACCACTATTTACCACTATTCGTACTATGGGAGTATTA 240

Qy 1084 ttgagctactgacattttgtgattggtggaattgccatgtctttttcacatggatc 1143
Db 239 TTCACCAAGCAATGGATCTCTCTTTTGGTAACACATTTATTGGCATTACACAAGTGGATAT 180

Qy 1144 ctgagcgctctggaatgggatacactccaaacgctgtgcacatctcactactcaagattt 1203
Db 179 TTTCTCATCTTTTGGGAATGATGATATACACCAAGCTGTCTGCCACCGGAATACTCAAAATTA 120

Qy 1204 gccgctcagctttcgtttgcactctatggttggtgccttctcacccgtggc 1254
Db 119 GCTGGTCAAGTGTCTGCACTGCTCTCTGTTCTTGTTACTGCTGTGTC 69

RESULT 10
AUI15402/c      300 bp mRNA EST 19-OCT-2000
LOCUS
DEFINITION      AUI15402 unpublished oligo-capped cDNA library Caenorhabditis
elegans cDNA clone yk732c11 3', mRNA sequence.
ACCESSION      AUI15402
VERSION
KEYWORDS
SOURCE
ORGANISM        Caenorhabditis elegans.
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
AUTHORS        Kohara,Y., Shin-i.T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE          A complementary view of the C. elegans genome
JOURNAL
COMMENT        Unpublished (2000)
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
FEATURES
source
1..300
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk732c11"
/clone_lib="unpublished oligo-capped cDNA library"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT      93 a 68 c 56 g 83 t
ORIGIN

Query Match      4.1%; Score 57.6; DB 10; Length 300;
Best Local Similarity 52.5%; Pred. No. 0.072;
Matches 126; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy 1022 ttcttcattctctctctgaactatcggtgcacagcgctgttaccctgttttc 1081
Db 300 TATTCATTCCATTATTTATCTTTTCTGTAATATGCTCCAGATAGTAGGATGGCCTGTTT 241

Qy 1082 tcttgagctactgacattttgtgattggtggaattgccatgtctttttcacatgat 1141
Db 240 GGTTTAAGAAATGAATGGTGGTTCACATTATGCAATTAATGGCAATTTACTTGTGTT 181

Qy 1142 acctcagcgctctggaatgggatacactccaaacgctgtgcacatctcactactcaagat 1201
Db 180 ATATGACAGCTTTAGCGCTGATTTATACACCAAGCAAGTGGCCAGCTAGATATCAGAAGC 121

Qy 1202 ttgccgctcagctttccggtttgcaactcttattggttgccctctcaccggtggc 1261
```

```
Qy 1015 cgtgctttttcattccattcttcttcttgcactatcggtgcacagcgctgttacc 1074
Db 299 CGAACCGTATTCATTCCTATTTATCTTTCTGTAATATCTGTCAGATAGTAGAAGTGG 240

Qy 1075 cctgtttttttgagctactgacattttgtgattggtggaattgccatgtctttttca 1134
Db 239 CCTGTTTGGTTTAAAGAAATGAATGGTGGTTCACCTATTGGATGTACCAATAATGGCAATTTACT 180

Qy 1135 catgataacctcagcgctctggaatgggatacactccaaacgctgtgcacatctcactac 1194
Db 179 TGTGGTTTATATGACGACTTTAGCGCTGATTTATACACCAAGCAAGTGGCCAGCTAGATAT 120

Qy 1195 tcaagatttgcgctcagctctccgttttgcactcttattggttggtccttctcacccgtggc 1254
Db 119 CAGAAGCTAAGTGAATGCTTGCATCAATTTCTCTAATGCTCGGAATTCATCGGAGTC 60

RESULT 11
AUI14995/c      300 bp mRNA EST 19-OCT-2000
LOCUS
DEFINITION      AUI14995 unpublished oligo-capped cDNA library Caenorhabditis
elegans cDNA clone yk724f10 3', mRNA sequence.
ACCESSION      AUI14995
VERSION
KEYWORDS
SOURCE
ORGANISM        Caenorhabditis elegans.
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
AUTHORS        Kohara,Y., Shin-i.T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE          A complementary view of the C. elegans genome
JOURNAL
COMMENT        Unpublished (2000)
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
FEATURES
source
1..300
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk724f10"
/clone_lib="unpublished oligo-capped cDNA library"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT      96 a 67 c 53 g 84 t
ORIGIN

Query Match      4.1%; Score 57.4; DB 10; Length 300;
Best Local Similarity 49.5%; Pred. No. 0.079;
Matches 148; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

Qy 1022 ttcttcattctctctctgaactatcggtgcacagcgctgttaccctgttttc 1081
Db 300 TATTCATTCCATTATTTATCTTTTCTGTAATATGCTCCAGATAGTAGAAGTGGCCTGTTT 241

Qy 1082 tcttgagctactgacattttgtgattggtggaattgccatgtctttttcacatgat 1141
Db 240 GGTTTAAGAAATGAATGGTGGTTCACATTATGCAATTAATGGCAATTTACTTGTGTT 181

Qy 1142 acctcagcgctctggaatgggatacactccaaacgctgtgcacatctcactactcaagat 1201
Db 180 ATATGACAGCTTTAGCGCTGATTTATACACCAAGCAAGTGGCCAGCTAGATATCAGAAGC 121

Qy 1202 ttgccgctcagctttccggtttgcaactcttattggttgccctctcaccggtggc 1261
```

```

Db 120 TAAGTGAATGCTGCATCAATTTCCCTAATGCTCGGAATTCATCGGAGTCGCCAGTA 61
Qy 1262 ccgtgtgtattgagcaacttcggtggacaagccaagtatcttataaataattatagatta 1320
Db 60 CACCATTGCTGCATGGCGGTGGAGTCTGATAGGAAGTAGAAGAACATGACTGTATAA 2

RESULT 12
CNS005TE 997 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL060767
KEYWORDS AL060767.1 GI:4943573
SOURCE GSS.
ORGANISM fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 997)
REFERENCE
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1..997
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR12K22"
/note="end : TET3"
BASE COUNT 89 a 99 c 13 g 258 t 538 others
ORIGIN

Query Match 4.0%; Score 56.2; DB 13; Length 997;
Best Local Similarity 18.6%; Pred. NO. 0.1;
Matches 116; Conservative 172; Mismatches 335; Indels 0; Gaps 0;

Qy 661 acaagcaagatttaccacattaccatcaaaaaggaaattcggaagcgaagcgcg 720
Db 356 ANNNAANNNAANNNAANNNAANNNTNANNNAANNAANNAANNNTATGNNNNN 415
Qy 721 gaaacgcagacgcgtctccattcttggaccacattcacaaactgttatggcga 780
Db 416 NNNNNNNNAANNNAANNNAANNANGCTNNNNNNNNNNNNNNNNNNNNNNNNNN 475
Qy 781 ctcttcaatgttgcgttcgttcggttactctacaaatctccctgttatgatgacc 840
Db 476 TTTTGTTTTGTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTT 535
Qy 841 gttaccactcgtggagatcccggtctctcacaacaaatattctcgaacacgatgaa 900
Db 536 TCTCTCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTT 595

```

```

Qy 901 tacatttgctcacaagtttctctcgtcttcaattgttcgtcgattgagaccatagtt 960
Db 596 CYTCTTTCYCYCYCTTTCYCYCYCTTTCYCYCYCTTTCYCYCYCTTTCYCYCYCTT 555
Qy 961 gcttccaagattcactgcccacaccgcgtttacctcaaatattgccaataatcttgctgt 1020
Db 656 YVAYKVCYCYCYCTTTCYCYCYCTTTCYCYCYCTTTCYCYCYCTTTCYCYCYCTT 715
Qy 1021 cttttcattccattcttcttcttctgcaactatcgtgtccacagacgcgtcttactcgtt 1080
Db 716 YFTCTTCYCYCYCTTTCYCYCYCTTTCYCYCYCTTTCYCYCYCTTTCYCYCYCTT 775
Qy 1081 tctcttgagctactgacattttgttgatggatgggaatgcatctcttttccacatgga 1140
Db 776 YVCCYCYCYCTTTCYCYCYCTTTCYCYCYCTTTCYCYCYCTTTCYCYCYCTT 835
Qy 1141 tacctcagcgtctgccaatgggatacactccaacacgtcgtgccatctcactactaca 1200
Db 836 YCTVCTCTTCYCYCTTTCYCYCYCTTTCYCYCYCTTTCYCYCYCTTTCYCYCYCTT 895
Qy 1201 ttgctcgtcagcttctcgttttgcactcttattggttggccttctcaccggttgctgtg 1260
Db 896 TTYCCYCYCYCTTTCYCYCYCTTTCYCYCYCTTTCYCYCYCTTTCYCYCYCTT 955
Qy 1261 ccggttgattgagcacttcgt 1283
Db 956 TTYTYTYTYTYCTTCYCYCYCTTTCYCYCYCTTTCYCYCYCTTTCYCYCYCTT 978

RESULT 13
AA406898 258 bp mRNA EST 01-MAY-1997
LOCUS MBACFZ7F07T3 Brugia malayi adult female cDNA (SAW96MLW-BmaF) Brugia
DEFINITION malayi cDNA clone AFCZ7F07 5', mRNA sequence.
ACCESSION AA406898
VERSION AA406898.1 GI:2064981
KEYWORDS EST.
SOURCE Brugia malayi.
ORGANISM Brugia malayi.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Brugia.
REFERENCE 1 (bases 1 to 258)
AUTHORS Blaxter,M.L., Waterfall,M., Daub,J., Lizotte,M., Baron,L. and Jones
,S.J.
TITLE Genes expressed in adult female Brugia malayi
JOURNAL Unpublished (1996)
COMMENT Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The ABI trace of this sequence can be viewed at
http://www.sanger.ac.uk/brugia/AFC/MBACFZ7F07T3.html
Seq primer: T3
Location/Qualifiers
1..258
/organism="Brugia malayi"
/db_xref="taxon:6279"
/clone="AFCZ7F07"
/clone_lib="Brugia malayi adult female cDNA (SAW96MLW-BmaF)"
/sex="female"
/dev_stage="adult"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Lymphatic filarial nematode parasite of humans.
mRNA was prepared from approximately 50 adult females
isolated from the peritoneal cavity of jirds and
converted to double-stranded cDNA using reverse

```



Tel: 81-559-81-6854  
Fax: 81-559-81-6855  
Email: ykohara@lab.nig.ac.jp.  
Location/Qualifiers

Location/Qualifiers

1. .345

/organism="Caenorhabditis elegans"

```

/origin="C
/strain="N2"

```

```
/db xref="taxon:6239"
```

```
/usr_xr1= caxon:
/clone="vk709a4"
```

```
/clone= yk703a4
/clone_lib="unpublished oligo-capped cDNA library"
```

```
/sex="Hermaphrodite"
```

/tissue\_type="whole animal"

```
/dev stage="varied"
```

/dev_stage=	varied		
86 c	67 q	91 t	1 others

ORIGIN

Query Match	3.6%;	Score 50.4;	DB 10;	Length 345;
Best Local Similarity	69.0%;	Pred. No. 2;		
Matches 69;	Conservative 0;	Mismatches 31;	Indels 0;	Gaps 0;

Qy 52 ccaagagacaagtacaatatattgtctactggctcgtcattcttqttggattcggagttctt 111

Db 106 CCAGAGGACAAGGGAAATCTTGTTTTCTACATATATCTTGCTCCATGGAAATCGGAACTCTT 165

Qy 112 ctgccatggaatatgttctattactatcgccctgagtatt 151

Db 166 ATGCCATGGAACATGCTTATTACGATCTCCTATGATTATT 205

Search completed: February 28, 2002, 00:25:49  
Job time: 3830 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2002, 23:24:08 ; Search time 49.03 Seconds  
(without alignments)  
6411.411 Million cell updates/sec

Title: US-08-816-011f-36  
Perfect score: 1388  
Sequence: 1 atggtaataatcaacgacg.....ttatttaaaaaaaaaaaaaa 1388

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Missing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	4.9	7218	1	US-08-232-463-14
2	38.6	2.8	1582	3	US-08-545-196B-10
3	38.6	2.8	1582	3	US-08-545-196B-12
4	36	2.6	1164	2	US-08-794-796-1
5	36	2.6	2621	2	US-08-553-619B-8
6	36	2.6	5852	1	US-07-867-106-2
7	35.8	2.6	1400	1	US-08-464-164-1
8	35.8	2.6	1400	1	US-08-338-057-1
9	35.8	2.6	1400	2	US-08-668-416-1
10	35.4	2.6	581	3	US-08-557-309B-22
11	35.4	2.6	581	3	US-08-834-306-22
12	35.4	2.6	581	4	US-08-993-674A-22
13	35.4	2.6	1303	2	US-08-793-410-11
14	35.4	2.6	3489	2	US-08-728-323A-1
15	35.4	2.6	32207	2	US-08-770-379-20
16	35.4	2.6	32207	4	US-08-757-669A-20
17	35.2	2.5	972	3	US-09-286-690-1
18	35	2.5	51259	3	US-08-781-891-209
19	34.8	2.5	289	4	US-09-007-005-17
20	34.8	2.5	289	4	US-09-244-796-17
21	34.6	2.5	991	4	US-09-377-648-7
22	34.6	2.5	2238	1	US-08-742-011-1
23	34.6	2.5	3645	2	US-08-663-112-1
24	34.2	2.5	852	4	US-09-461-697-1
25	34.2	2.5	1991	2	US-08-415-593-40
26	34.2	2.5	2128	2	US-08-415-593-39
27	33.8	2.4	519	1	US-08-339-582-1

28	33.6	2.4	732	4	US-08-916-576B-5	Sequence 5, Appli
29	33.4	2.4	756	2	US-08-530-165-1	Sequence 1, Appli
30	33.2	2.4	174	2	US-08-378-235B-8	Sequence 8, Appli
31	33.2	2.4	248	4	US-09-007-005-32	Sequence 32, Appli
32	33.2	2.4	248	4	US-09-244-796-32	Sequence 32, Appli
33	33.2	2.4	277	4	US-09-007-005-3	Sequence 3, Appli
34	33.2	2.4	277	4	US-09-244-796-3	Sequence 3, Appli
35	33.2	2.4	1364	1	US-08-265-087-3	Sequence 3, Appli
36	33.2	2.4	1364	1	US-08-621-493-3	Sequence 3, Appli
37	33.2	2.4	1364	2	US-08-965-688-3	Sequence 3, Appli
38	33.2	2.4	1364	4	US-09-260-173-3	Sequence 3, Appli
39	33.2	2.4	2265	3	US-09-369-618-3	Sequence 3, Appli
40	33.2	2.4	2265	3	US-09-369-617-3	Sequence 3, Appli
41	33	2.4	306	4	US-09-122-400B-9	Sequence 9, Appli
42	33	2.4	1939	1	US-07-715-751B-2	Sequence 2, Appli
43	33	2.4	3498	3	US-08-293-728-1	Sequence 1, Appli
44	33	2.4	3498	4	US-09-421-868-1	Sequence 1, Appli
45	33	2.4	5852	1	US-07-867-106-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-Fls  
US-08-232-463-14

```

Query Match      4.9%; Score 68; DB 1; Length 7218;
Best Local Similarity 4.2%; Pred. No. 1.4e-09;
Matches 17; Conservative 234; Mismatches 149; Indels 0; Gaps 0;

QY 907 ttgctcaagatttctcgtcttcaatttggctgcgagattggatccatagttgcttc 966
Db 1060 TTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1119
QY 967 aagattcactggcgacaccccgcttacctcaaatgtgccaatacttgctgctcttc 1026
Db 1120 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1179
QY 1027 attcattcttcttcgtcaactatgctgcagacgctgcttctcttcttctt 1086
Db 1180 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1239
QY 1087 gactcactgacattttgtgattggtggaattggcattgctcttttcacatgatacctc 1146
Db 1240 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1299
QY 1147 agcgtctggcaatgggatacacactccaacgctgctgccaatcactactcaagattgcc 1206
Db 1300 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1359
QY 1207 gctcagcttccgttgacactcttattggtggccttcacccgctggtgcccgtt 1266
Db 1360 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1419
QY 1267 gttattgacactctgagacagcgaagtattcttataa 1306
Db 1420 YYYYYYYYYYYYYYATACCAAAATCTCTATCTCTTTAA 1459

```

```

RESULT 2
US-08-545-196B-10
; Sequence 10, Application US/08545196B
; Patent No. 6080577
; GENERAL INFORMATION:
; APPLICANT: MELKI, JUDITH
; APPLICANT: MUNNICH, ARNOLD
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA

```

```

US-08-545-196B-10
Query Match      2.8%; Score 38.6; DB 3; Length 1582;
Best Local Similarity 63.4%; Pred. No. 0.17;
Matches 59; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1296 tatcttataaattatttagcattagctatctgttatctgttatttatttaagctg 1355
Db 1441 TATCTTCTATATGTTTAAAGTATATAATAAATAATTTTATTTTAAAAA 1500
QY 1356 tggaataaataattattataaaaaa 1388
Db 1501 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1533

```

```

RESULT 3
US-08-545-196B-12
; Sequence 12, Application US/08545196B
; Patent No. 6080577
; GENERAL INFORMATION:
; APPLICANT: MELKI, JUDITH
; APPLICANT: MUNNICH, ARNOLD
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-545-196B-12
Query Match      2.8%; Score 38.6; DB 3; Length 1582;
Best Local Similarity 63.4%; Pred. No. 0.17;
Matches 59; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1296 tatcttataaattatttagcattagctatctgttatctgttatttatttaagctg 1355
Db 1441 TATCTTCTATATGTTTAAAGTATATAATAAATAATTTTATTTTAAAAA 1500
QY 1356 tggaataaataattattataaaaaa 1388
Db 1501 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1533

```





TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-568-3100  
TELEFAX: 215-568-3439  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5852 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2378..5038  
NAME/KEY: CDS  
LOCATION: 2378..5038  
US-07-867-106-2

Query Match 2.6%; Score 36; DB 1; Length 5852;  
Best Local Similarity 55.6%; Pred. No. 1.8;  
Matches 69; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
QY 1265 ttgtattgagcacttcgtgacagccagctatcttataataatttatagcattagagt 1324  
Db 2096 TTATTAATAAAAAAATCAAAAAAACCAGTAATATTTATATATGAGGGTTTTTT 2037  
QY 1325 atacttgattatgttcttttattagctgtggaataataattattataaaaaaaa 1384  
Db 2036 TTTTCTTTTTTTTTTTTTTTTCAAGTAAAAAATAAAAAAAGAAATA 1977  
QY 1385 aaaa 1388  
Db 1976 GAAA 1973

RESULT 7  
US-08-464-164-1  
Sequence 1, Application US/08464164  
Patent No. 5614195  
GENERAL INFORMATION:  
APPLICANT: Tomley, Fiona M.  
APPLICANT: Dunn, Paul P. J.  
APPLICANT: Bumstead, Janene M.  
APPLICANT: Vermeulen, Arno N.  
TITLE OF INVENTION: Coccidiosis poultry vaccine  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Akzo No. 5614195el Patent Department  
STREET: 1300 Piccard Drive, Suite 206  
CITY: Rockville  
STATE: Maryland  
COUNTRY: U.S.A.  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/464,164  
FILING DATE: June 2, 1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gormley, Mary E.  
REGISTRATION NUMBER: 34,409  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 258-5200  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1400 base pairs  
TYPE: nucleic acid

STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: C-terminal  
ORIGINAL SOURCE:  
ORGANISM: Eimeria maxima  
STRAIN: Houghton  
DEVELOPMENTAL STAGE: sporozoite  
IMMEDIATE SOURCE:  
LIBRARY: sporozoite cDNA cloned in Lambda ZAPII  
CLONE: Em70-1  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1368  
US-08-464-164-1

Query Match 2.6%; Score 35.8; DB 1; Length 1400;  
Best Local Similarity 69.0%; Pred. No. 1;  
Matches 49; Conservative 0; Mismatches 22; Indels 0; Gaps 0;  
QY 1318 tttagtgatacttggtatgtgtttttatttaagctgtggaataataattattataaa 1377  
Db 1326 TCAGCAAAATGCTTCTTAATATTATGTGTAATCTGCAGCAGATAATAATAATAATAA 1385  
QY 1378 aaaaaaaaaa 1388  
Db 1386 AAAAAAAAAA 1396

RESULT 8  
US-08-338-057-1  
Sequence 1, Application US/08338057  
Patent No. 5795741  
GENERAL INFORMATION:  
APPLICANT: Tomley, Fiona M.  
APPLICANT: Dunn, Paul P. J.  
APPLICANT: Bumstead, Janene M.  
APPLICANT: Vermeulen, Arno N.  
TITLE OF INVENTION: Coccidiosis poultry vaccine  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Organon Teknika Corporation  
STREET: 1330 Piccard Drive  
CITY: Rockville  
STATE: Maryland  
COUNTRY: U.S.A.  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/338,057  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 93.309078.9  
FILING DATE: 12-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Gormley, Mary E.  
REGISTRATION NUMBER: 34,409  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 258-5200  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1400 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

Query Match	2.6%	Score 35.8;	DB 2;	Length 1400;
Best Local Similarity	69.0%;	Pred. No. 1;		

Query Match 2.6%; Score 35.4; DB 2; Length 581;  
Best Local Similarity 61.3%; Pred. No. 0.83;



```

; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,410
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10627
; FILING DATE: AUGUST 25, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CHRISTENBURY, LYNNE M.
; REGISTRATION NUMBER: 30,971
; REFERENCE/DOCKET NUMBER: CR-9567-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-5481
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-793-410-11

Query Match 2.6%; Score 35.4; DB 2; Length 1303;
Best Local Similarity 61.3%; Pred. No. 1.3;
Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1295 gtatcttataaatttatagcattagtagtactgtgtatattgtttttataagct 1354
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 639 GTAATATATATTTTGTATATATATATATATATATATATATATATATATATAT 580

QY 1355 gtggaataaataattattcaaaaaa 1387
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 579 ATTGTTAGATTTCACCTGAACAAAAAAA 547

RESULT 14
US-08-728-323A-1/C
; Sequence 1, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728.323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3207 base pairs
; TYPE: nucleic acid

; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,410
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10627
; FILING DATE: AUGUST 25, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CHRISTENBURY, LYNNE M.
; REGISTRATION NUMBER: 30,971
; REFERENCE/DOCKET NUMBER: CR-9567-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-5481
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-793-410-11

Query Match 2.6%; Score 35.4; DB 2; Length 3489;
Best Local Similarity 57.8%; Pred. No. 2.1;
Matches 63; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 264 cctgttcctcattattgtgtccctcctatctaccgcgtcttcttcggtttgtctcaa 323
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1282 CCTCCTCGTCCCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1223

QY 324 catcgtaaacctgacaaatctctctctctctctctctctctctctctctctctctct 372
||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1222 CATCGTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1174

RESULT 15
US-08-770-379-20
; Sequence 20, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3207 base pairs
; TYPE: nucleic acid

```

```
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-770-379-20
```

```
Query Match      2.6%; Score 35.4; DB 2; Length 32207;
Best Local Similarity 57.8%; Pred. No. 6.3;
Matches 63; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 264 cctgttctcattattgctggtccctgatctaccgcggtctttgtctcggtttgcttcaa 323
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20715 CCTCGTCTCTCTCTTGTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAT 20774
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 324 catcgtcaacctgacaaatcatctcctcgtcgtcattgttctgagccc 372
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20775 CATCGTCATCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 20823
```

```
Search completed: February 28, 2002, 00:50:28
Job time: 5180 sec
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 16:59:34 ; Search time 13.23 seconds  
(without alignments)  
1202.762 Million cell updates/sec

Title: US-08-816-011f-63

Perfect score: 2294

Sequence: 1 MVIINRSNTYAVQAEAPRD.....LTGGLMPVIEHFVDRKPSIL 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %	Length	DB	ID	Description
1	343.5	15.0	455	1	ENT1_HUMAN	Q99808 homo sapien
2	326.5	14.2	456	1	ENT2_HUMAN	Q14542 h equilibra
3	320.5	14.0	456	1	ENT2_RAT	O54699 rattus norv
4	301	13.1	456	1	ENT1_RAT	O54698 rattus norv
5	209.5	9.1	327	1	ENT2_MOUSE	O61672 m equilibra
6	182.5	8.0	517	1	FU26_YEAST	P31381 saccharomyc
7	134	5.8	432	1	Y672_METUA	Q58086 methanococc
8	124.5	5.4	506	1	HITB_HAEIN	P71338 haemophilus
9	117.5	5.1	559	1	YABC_SCHPO	Q09852 schizosacch
10	113.5	4.9	477	1	Y098_MYCGE	P47344 mycoplasma
11	113.5	4.9	587	1	T9S3_MOUSE	Q9et30 mus musculu
12	113.5	4.9	589	1	T9S3_HUMAN	Q9rd45 homo sapien
13	113.5	4.9	604	1	Y05M_HORSE	P48656 equus cabal
14	113	4.9	459	1	TCR2_BACSU	P14512 bacillus su
15	113	4.9	614	1	YDKN_LACLC	P42377 lactococcus
16	112	4.9	383	1	NAPK_ENTHR	P26235 enterococcu
17	111.5	4.9	397	1	PAR2_HUMAN	P55085 homo sapien
18	111.5	4.9	433	1	TCR2_STAAU	P02983 staphylococ
19	111.5	4.9	439	1	UHPG_ECOLI	P09836 escherichia
20	110.5	4.8	479	1	Y098_MYCPN	P75535 mycoplasma
21	110.5	4.8	785	1	ISP4_SCHPO	P40900 schizosacch
22	108.5	4.7	442	1	UHPG_SALTY	P27669 salmonella
23	108	4.7	388	1	YUBA_BACSU	O32086 bacillus su
24	106.5	4.6	399	1	PAR2_MOUSE	P55086 mus musculu
25	106	4.6	547	1	Y05M_ASCSU	P24884 ascaris suu
26	105.5	4.6	598	1	THIX_YEAST	Q08485 saccharomyc
27	105	4.6	542	1	ATR1_YEAST	P13090 saccharomyc
28	104.5	4.6	894	1	YN86_YEAST	P27514 saccharomyc
29	104	4.5	527	1	NU2M_ACACA	Q37376 acanthamoeb
30	104	4.5	552	1	Y05M_RHISA	Q9zym7 rhinicephal
31	103.5	4.5	345	1	NUOH_RHOCA	P42032 rhodobacter
32	103.5	4.5	606	1	Y05M_EQUAS	P92485 equus asinu
33	103	4.5	459	1	NU4M_BOVIN	P03910 bos taurus

#### RESULT 1

ID	ENT1_HUMAN	STANDARD;	PRT;	455 AA.
AC	Q99808;			
DT	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	EQUILIBRATIVE NUCLEOSIDE TRANSPORTER 1 (EQUILIBRATIVE			
DE	NITROBENZYLMECAPTOPURINE RIBOSIDE-SENSITIVE NUCLEOSIDE TRANSPORTER)			
DE	(EQUILIBRATIVE NBMPR-SENSITIVE NUCLEOSIDE TRANSPORTER) (NUCLEOSIDE			
DE	TRANSPORTER, ES-TYPE).			
GN	SLC29A1 OR ENT1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-21.			
RC	TISSUE=Placenta;			
RX	MEDLINE=97140266; PubMed=8986748;			
RA	Griffiths M., Beaumont N., Yao S.Y.M., Sundaram M., Boumah C.E.,			
RA	Davies A., Kwong F.Y.P., Coe I., Cass C.E., Young J.D., Baldwin S.A.;			
RT	"Cloning of a human nucleoside transporter implicated in the cellular			
RT	uptake of adenosine and chemotherapeutic drugs.;"			
RL	Nat. Med. 3:89-93(1997).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Small intestine, and Jejunum;			
RA	Lum P.Y., Ngo L.Y., Bakken A.H., Unadkat J.D.;			
RT	"Critical structural determinants for high affinity binding of			
RT	nucleosides to the equilibrative NBMPR-sensitive nucleoside			
RT	transporter (es) cloned from the human jejunum.;"			
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.			
CC	!- FUNCTION: MEDIATES BOTH INFILUX AND EFFLUX OF NUCLEOSIDES ACROSS			
CC	THE MEMBRANE (EQUILIBRATIVE TRANSPORTER). IT IS SENSITIVE (ES) TO			
CC	LOW CONCENTRATIONS OF THE INHIBITOR NITROBENZYLMECAPTOPURINE			
CC	RIBOSIDE (NBMPR) AND IS SODIUM-INDEPENDENT. IT HAS A HIGHER			
CC	AFFINITY FOR ADENOSINE. INHIBITED BY DIPYRIDAMOLE AND DILAZEP			
CC	(ANTICANCER CHEMOTHERAPEUTICS DRUGS).			
CC	!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.			
CC	!- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, MAMMARY GLAND,			
CC	ERYTHROCYTES AND PLACENTA, AND ALSO IN FETAL LIVER AND SPLEEN.			
CC	!- PTM: GLYCOSYLATED.			
CC	!- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; U81375; AAC51103.1; -			
DR	EMBL; AF079117; AAC62495.1; -			
DR	MIM; 602193; -			

O51750 borrelia bu  
O37370 acanthamoeb  
O33814 staphylococ  
O60779 homo sapien  
O46522 b cytochrom  
P49219 xenopus lae  
P39637 bacillus su  
P39301 escherichia  
P04371 trypanosoma  
P57543 buchnera ap  
Q46916 escherichia  
P25568 saccharomyc

#### ALIGNMENTS







Db 67 NNWYTLSSQLPPLLFTLLNSFLYQCIPESVRILG-----SLIAILLFALTAALVKVDLSP 122  
Qy 130 SWFFWVTLGMAISINFSNGLYENSIVGVGGDFPHYTGALLIGNNICGL-----LIIVVK 184  
Db 123 GLFFSITMASVWFVNSCAVLGSLFOLGTMPTSTYTLFSLGGGLAGIFAALAMLTSLA 182  
Qy 185 IGVTFYFLNDPKLVAIVVF-----GISLVLLVCAIAL-----FFITKQ----- 223  
Db 183 SGV-----DPQTSALGYFITPCVGILLISICYLSPHLKFAFYLLTKKQAPVQLETK 236  
Qy 224 -DFVYHHQKMEIR-----EK-----AETDRSPSILMTTTCYVGLFN 264  
Db 237 AELLGADEKNGIPVSPQOAGPTLDLDPKEKELEGLEEPQPKPSVFWFKIWLALCL 296  
Qy 265 WFCFAVTLTIFP-VMMVTVTTRGDSGLINKIMSENDEIYTLTSLFVNLFAAGISIVASK 323  
Db 297 VLVTFTVLSVFPALTAMVTSSNS-----PGKWSQFFNPICCELLFNVDWGLRSLTSY 350  
Qy 324 IHWP--TPRYLKFAIILRALFIPFFFCNRYRVOTRAYVPVFESDIFVIGSIAMFSHG 381  
Db 351 FLWPDDESQLLPLLVCLRFLEVPFLMCHVPQAR-LPIIFWQDAYFITPMLLFAISNGY 409  
Qy 382 LSLAMGYTPNVVPSHYSRFAAOLSVCTLMVGLLTG 417  
Db 410 FVSLTMCAPRQVLPHEREVAGALMTFFLALGLSCG 445

## RESULT 4

ENTL\_RAT ID ENTL\_RAT STANDARD; PRT; 456 AA.  
AC 054698;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE EQUILIBRATIVE NUCLEOSIDE TRANSPORTER 1 (EQUILIBRATIVE  
DE NITROBENZYL MERCAPTOPYRINE RIBOSIDE-SENSITIVE NUCLEOSIDE TRANSPORTER)  
DE (EQUILIBRATIVE NEMPR-SENSITIVE NUCLEOSIDE TRANSPORTER) (NUCLEOSIDE  
DE TRANSPORTER, ES-TYPE)  
GN SLC29A1 OR ENTL1  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxId=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPRAIN=SPRAGUE-DAWLEY; TISSUE=Jejunum;  
RX MEDLINE=98019212; PubMed=933301;  
RA Yao S.Y.M., Ng A.M.L., Muzyka W.R., Griffiths M., Cass C.E.,  
RA Baldwin S.A., Young J.D.;  
RT "Molecular cloning and functional characterization of  
RT nitrobenzylchinosine (NEMPR)-sensitive (es) and NEMPR-insensitive  
RT (ei) equilibrative nucleoside transporter proteins (rent1 and rent2)  
RT from rat tissues.";  
RL J. Biol. Chem. 272:28423-28430(1997).  
CC -!- FUNCTION: MEDIATES BOTH INFLUX AND EFFLUX OF NUCLEOSIDES ACROSS  
CC THE MEMBRANE (EQUILIBRATIVE TRANSPORTER). IT IS SENSITIVE (ES) TO  
CC LOW CONCENTRATIONS OF THE INHIBITOR NITROBENZYL MERCAPTOPYRINE  
CC RIBOSIDE (NEMPR) AND IS SODIUM-INDEPENDENT. IT HAS A HIGHER  
CC AFFINITY FOR ADENOSINE, RESISTANT TO DIPHYRIDAMOLE AND DILAZEP  
CC INHIBITION (ANTICANCER CHEMOTHERAPEUTICS DRUGS).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN JEJENUM, LIVER AND LUNG.  
CC -!- PTM: GLYCOSYLATED.  
CC -!- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF015304; AAB88049.1; --  
DR InterPro; IPR002259; DER\_Nucleoside\_tran.  
DR Pfam; PF01733; Nucleoside\_tran; 2;  
DR PRINTS; PR01130; DERENTRNSPRT.  
DR ProDom; PD005103; DER\_Nucleoside\_tran; 1.  
KW Transmembrane; Transport; Glycoprotein.  
FT INIT\_MET 0 0  
FT DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 12 28 POTENTIAL.  
FT DOMAIN 29 81 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 82 106 POTENTIAL.  
FT DOMAIN 107 110 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 111 129 POTENTIAL.  
FT DOMAIN 130 137 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 138 156 POTENTIAL.  
FT DOMAIN 157 173 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 174 198 POTENTIAL.  
FT DOMAIN 199 205 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 206 226 POTENTIAL.  
FT DOMAIN 227 290 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 291 310 POTENTIAL.  
FT DOMAIN 311 322 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 323 342 POTENTIAL.  
FT DOMAIN 343 359 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 360 378 POTENTIAL.  
FT DOMAIN 379 393 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 394 413 POTENTIAL.  
FT DOMAIN 414 431 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 432 452 POTENTIAL.  
FT DOMAIN 453 456 EXTRACELLULAR (POTENTIAL).  
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 456 AA; 49885 MW; A34CE92C20836D9B CRC64;

Query Match. 13.1%; Score 301; DB 1; Length 456;  
Best Local Similarity 22.3%; Pred. No. 6.2e-14;  
Matches 107; Conservative 84; Mismatches 191; Indels 98; Gaps 17;

Qy 18 PRKYNIVYVNLVILVGVGLLPNNMTTIAPEYVYVWFKPDGV-----ETWYSKEPMGS 72  
Db 6 PDORYKAVVLIFFVLGLTLLPNWFFIT-ATQYFTRSLNTSQNISLVTSNCSCESTALAD 64  
Qy 73 LTIGSQLPNASINVFNLFIIAGPLIYRVAPVCFN-----IVNLTII 115  
Db 65 PSVSLPARSSLSAIFNNVWTLCAMLPLLI--TCLNSFLHKKVQSQRILGSLALLVF 122  
Qy 116 LILVIVLEPTEDSMWFFWVTLGMAISINFSNGLYENSIVGVGGDFPHYTGALLIGNNI 175  
Db 123 LVTATLVKQVMDALS-FFIITMIKIVLINSFGAILQASLFGLAGVLPANVTAPIMSGQL 181  
Qy 176 CGLLITVVKI-----GVTFPLNDPKLVAIVVFGISLVLLVCAIALFFITKQ 223  
Db 182 AGFTSVAMICAVASGSKLSAAGYFVIT---ACAVV-----ILAICYLALPWM--- 228  
Qy 224 DFVYHHQ-----KGMETR-EKAETDRSPSTL----- 250  
Db 229 EYRHYQLNLNAGPAOETKLDLISEGEEPRGGRSESGVPGPNLSPPANRQSTKAIUKSI 288  
Qy 251 WTTFTTCYGLFNWFCFAVTLTIFPVMVTVTTRGDSGLINKIMSENDEIYTLTSLFV 310  
Db 289 WVL-----ALSVCFITFTVIGLFP---AVTAEVSSIAGTSPWKNK-FEIPVACFLNF 337  
Qy 311 NLFAAIGTSIVASIKHP--TPRYLKFAIILRALFIPFFFCNRYRVOTRAYVPVFESDIF 368  
Db 338 NVFDWGLRSLTAICMPPGQDSRWLPVLVACRVVFIPILLMLCNVK-QHHYLPSPFKHDWVF 396  
Qy 369 VIGGIAMFSHGYSALAMGYTPNVVPSHYSRFAAOLSVCTLMVGLLTGGLWPVIEHFV 428  
Db 397 ITFMAAFASNGYLSLCLMCFGPKVKPAPAEYAGNIMSFLLCLGLALGAVLSFLRALV 456

```

RESULT 5
ENT2_MOUSE
ID ENT2_MOUSE STANDARD; PRT; 327 AA.
AC Q61672;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EQUIILBRATIVE NUCLEOSIDE TRANSPORTER 2 (EQUILBRATIVE
DE NITROBENZYLPERCAPTOPURINE RIBOSIDE-INSENSITIVE NUCLEOSIDE TRANSPORTER)
DE (EQUILBRATIVE NMMP-INSENSITIVE NUCLEOSIDE TRANSPORTER) (NUCLEOSIDE
DE TRANSPORTER, EI-TYPE) (36 KDA NUCLEOLAR PROTEIN HNP36) (HYDROPHOBIC
DE NUCLEOLAR PROTEIN, 36 KDA) (DELAYED-EARLY RESPONSE PROTEIN 12).
GN SLC29A2 OR ENT2 OR HNP36 OR DER12.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=BALB/C; TISSUE=Fibroblast;
RX MEDLINE=95367016; PubMed=7639753;
RA Williams J.B., Lanahan A.A.;
RT "A mammalian delayed-early response gene encodes HNP36, a novel,
RT conserved nucleolar protein.";
RL Biochem. Biophys. Res. Commun. 213:325-333(1995).
CC -1- FUNCTION: MEDIATES EQUIILBRATIVE TRANSPORT OF PURINE AND
CC PYRIMIDINE NUCLEOSIDES, AND THE PURINE BASE HYPOXANTHINE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLAR. INTEGRAL MEMBRANE
CC PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM AND A SHORT
CC FORM/HNP36 (SHOWN HERE); SEEMS TO BE PRODUCED BY ALTERNATIVE
CC SPLICING.
CC -1- INDUCTION: BY PLATELET DERIVED GROWTH FACTOR (PDGF) AND FIBROBLAST
CC GROWTH FACTOR (FGF).
CC -1- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X86682; CAA60381.1; -
CC MGD; MGI:1345278; SLC29a2.
CC InterPro: IPR002259; DER_Nucleoside_tran.
CC Pfam; PF01733; Nucleoside_tran; 1.
CC PRINTS; PR01130; DERENTRNSPRT.
CC ProDom; PD005103; DER_Nucleoside_tran; 1.
CC KW Nuclear protein; Transmembrane; Transport; Alternative splicing.
CC FT TRANSMEM 2 22 POTENTIAL.
CC TRANSMEM 32 52 POTENTIAL.
CC TRANSMEM 63 83 POTENTIAL.
CC TRANSMEM 159 179 POTENTIAL.
CC TRANSMEM 194 214 POTENTIAL.
CC TRANSMEM 231 251 POTENTIAL.
CC TRANSMEM 267 287 POTENTIAL.
CC TRANSMEM 303 323 POTENTIAL.
CC SEQUENCE 327 AA; 36113 MW; 5D2D3FF4BBD592B6 CRC64;

Query Match 9.1%; Score 209.5; DB 1; Length 327;
Best Local Similarity 22.4%; Pred. No. 8.5e-08;
Matches 76; Conservative 49; Mismatches 134; Indels 81; Gaps 12;

QY 131 WFFVFWTLGMATSIINFSNGLYNSVYGVGGDFPHYTGALLIGNNICGLLITVWKI----- 185
|| : : : : : | | | | | : : : : :
DB 5 WF-----INSFCAVGLSFGQLGTMPSTYTLFSLGQGLAGIFAALMLMSLAS 54
|| : : : : : | | | | | : : : : :
QY 186 GVTFLNDPKLVAIVF---GISIVLLVCAIAL-----FFI-----TK 222
|| : : : : : | | | | | : : : : :

```

```

Db 55 GV-----DAQTSALGYFITPCVIGILLSIVCYLSPLHLKFAFYLTLEKLSQAQPTQLETK 108
QY 223 QDFYHYHHQKGMEIR-----EKAETDRPSPSILWTTTNCYGVGFNVWF 266
Db 109 AELQADEKNGVPISPQASPTLDLDPEKEPEEPQKPSVVFVFKHWTALCLVL 168
QY 267 CFAVTLTIFFVM--MTVTTRGDSG----FLNKIMSENDEIYLLTSFLVFNLFAAIGSIV 320
Db 169 VFTVLSVFPAITAMVTTSSNSPGKWLFFNP1-----CCFLLFNVMDWLGRLS 217
QY 321 ASKIHWP---TPRYLKFAIIRALPIPEFFFCNRYRVQTRAYVPVFESTDIFVIGIAMS 377
Db 218 TSYFLWPDDESQQLPLLVCLRFLEVPFLMFLCHVPOHAR-LPIIFRQDAYFITFMLLFAV 276
QY 378 SHGYLSALAMGYTPNVPSHYSRFAAQLSVCTLMVGLLTG 417
Db 277 SNGYLVSTMTCLAPROVLPHEREVAGALMTFFLALGLSCG 316

RESULT 6
FU26_YEAST
ID FU26_YEAST STANDARD; PRT; 517 AA.
AC P31381;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NUCLEOSIDE TRANSPORTER FUN26.
GN FUN26 OR YAL022C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN SEQUENCE FROM N.A.
RP STRAIN=S288C / AB972;
RX MEDLINE=93209532; PubMed=845870;
RA Ouellette B.F.F., Clark M.W., Keng T., Storms R.K., Zhong W.W.,
RA Zeng B., Fortin N., Delaney S., Barton A.B., Kaback D.B., Bussey H.;
RT "Sequencing of chromosome I from Saccharomyces cerevisiae: analysis
RT of a 32 kb region between the LTEL and SPO7 genes.";
RL Genome 36:32-42(1993).
RN CHARACTERIZATION.
RP MEDLINE=20408929; PubMed=10827169;
RA Vickers M.F., Yao S.Y., Baldwin S.A., Young J.D., Cass C.E.;
RT "Nucleoside transporter proteins of Saccharomyces cerevisiae.
RT Demonstration of a transporter (FU11) with high uridine selectivity
RT in plasma membranes and a transporter (FUN26) with broad nucleoside
RT selectivity in intracellular membranes.";
RL J. Biol. Chem. 275:25931-25938(2000).
CC -1- FUNCTION: HAS BROAD NUCLEOSIDE SELECTIVITY (URIDINE, ADEMOSINE AND
CC CYTIDINE) AND MOST LIKELY FUNCTIONS TO TRANSPORT NUCLEOSIDES
CC ACROSS INTRACELLULAR MEMBRANES.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L05146; AAC04935.1; -
CC EMBL; L05027; AAA70393.1; -
CC PIR; S36712; S36712.
CC SGD; S0000020; FUN26.
CC InterPro: IPR002259; DER_Nucleoside_tran.
CC Pfam; PF01733; Nucleoside_tran; 1.
CC ProDom; PD005103; DER_Nucleoside_tran; 1.
CC Transmembrane; Transport.
CC TRANSMEM 76 POTENTIAL.
FT

```





This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: Z64354; CAA91247.1; -  
InterPro: IPR003662; sub. transpstr.  
Pfam: PF00083; sugar\_tr; 1.  
DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 1.  
DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; 1.  
KW Hypothetical protein; Phosphate transport; Transmembrane;  
KW Glycoprotein. 45 65 POTENTIAL.  
FT TRANSMEM 189 209 POTENTIAL.  
FT TRANSMEM 336 356 POTENTIAL.  
FT TRANSMEM 382 402 POTENTIAL.  
FT TRANSMEM 410 430 POTENTIAL.  
FT TRANSMEM 475 495 POTENTIAL.  
FT TRANSMEM 498 518 POTENTIAL.  
FT CARBOHYD 454 454 N-LINKED (GLCNAC...) (POTENTIAL).  
SQ SEQUENCE 559 AA; 62316 MW; 2E8958F86C2092E2 CRC64;

Query Match 5.1%; Score 117.5; DB 1; Length 559;  
Best Local Similarity 20.1%; Pred. No. 0.29;  
Matches 101; Conservative 73; Mismatches 175; Indels 153; Gaps 25;  
Qy 26 YWLVI-----LVGFG-----VLLPNNFI--TIAPEYVYVWFKPDGVETWYSKEFGSL 73  
Db 34 HWLGLTRKREKLMGACAGFFLDYDLFIINLSPIVEYLYWGLGCKGHPYSGING-- 91  
Qy 74 TIGSQLPNASINVENFLIITAGPLIYR-----VFAPVCENIVNLTILIVILE 123  
Db 92 -----LVNAANIGNVF-----GQLFGMGDFGKRVYKGMIVVIAITVLVIALPKSI 142  
Qy 124 PTE-DSMSWFF--WVTLGMATSNFNSGLYNSVGVGGDFPHTYI-----GA 168  
Db 143 PTPLGKMMWIFAWRWL-LGL-----GIGGDYPSATITSESLSRGT 185  
Qy 169 LLI-----GNITCGLLITVVKIGTVYFNDEP-----KLVAI--VFYGLSVILL 211  
Db 186 LLSIVFSFGFGTLAIVTILLACF-----EKPLNQRGEYTKLEGVWRQLQMLALVPAL 241  
Qy 212 VCAI-----ALFFITKDFY-----HYHHKGMWEIREKAET-- 242  
Db 242 LVLPRLTMKESKYSQSKALNKYTDNDTYIADDEPKKNQNVVEBKQINLTSSDSHP 301  
Qy 243 -----DRPSPSLMTFTNCGYGFNVWPCF-----AVTLTIFVMMVTVTRGDSGF 289  
Db 302 TSTEDFGDKRASTVPTSENTSGFIEYSQHHFKHLLATAVSWFLLDIAFYGVNLSQSVI 361  
Qy 290 LNKI--MSENDEIYTLTSLFVFNFAAGISIVASKIHWPT-----PYRLKFAILLRALFI 343  
Db 362 LKAIGFSSGKNEVHTLMRG-AIGNLLIAIAGYVPG--YWFVTLVLEKLRKWKIQLGLFI 418  
Qy 344 PFFFCNRYVQTRAYPVFFESTDFIVGG-----IAMSFSHGVLALAMGYTPNVVPSH 397  
Db 419 TGLMFA-----ILAGSDWTISTGGRFACFVIAQFFSNFGPNATTLTYPAEVEPAR 468  
Qy 398 YSRFAAOLSVCTLMVGLLTGGL 419  
Db 469 VRGTAHGLSALGCKGAILASL 490

RESULT 10  
Y098 MYCGE STANDARD: PRT; 477 AA.  
AC P47344; Q49509;  
DT 01-OCT-1996 (Rel. 34, Created)  
DY 01-OCT-1996 (Rel. 34, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL PROTEIN MG098.  
GN MG098  
OS Mycoplasma genitalium.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=96026346; PubMed=7569993;  
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,  
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,  
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
RA "The minimal gene complement of Mycoplasma genitalium.";  
RN Science 270:397-403(1995).  
[2]  
RP SEQUENCE OF 46-155 AND 278-382 FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=94075230; PubMed=8253680;  
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;  
RA "A survey of the Mycoplasma genitalium genome by using random  
RT sequencing.";  
RL J. Bacteriol. 175:7918-7930(1993).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: U39690; AAC71316.1; -  
DR EMBL; U01701; AAB01013.1; -  
DR EMBL; U01782; AAD12771.1; -  
DR TIGR; MG098;  
DR InterPro: IPR003837; Glu-trnAGln.  
DR Pfam: PF02686; Glu-trnAGln; 1.  
KW Hypothetical protein; Transmembrane; Complete proteome.  
FT TRANSMEM 23 43 POTENTIAL.  
FT TRANSMEM 61 81 POTENTIAL.  
FT TRANSMEM 108 128 POTENTIAL.  
FT TRANSMEM 131 151 POTENTIAL.  
FT TRANSMEM 171 191 POTENTIAL.  
FT TRANSMEM 228 248 POTENTIAL.  
FT TRANSMEM 285 305 POTENTIAL.  
FT TRANSMEM 326 346 POTENTIAL.  
FT CONFLICT 87 87 F -> V (IN REF. 2).  
FT CONFLICT 372 382 DMOSLMPNVI->EYAKLNYAQCY (IN REF. 2).  
SQ SEQUENCE 477 AA; 54045 MW; 3FCDD95A2C35684 CRC64;  
Query Match 4.9%; Score 113.5; DB 1; Length 477;  
Best Local Similarity 19.7%; Pred. No. 0.47;  
Matches 89; Conservative 70; Mismatches 130; Indels 163; Gaps 20;  
Qy 22 YNIVYWLIVLVGFGVLLPNNFIITIAPEYVYVWFKPDGVETWYSKEF-----MGSL 73  
Db 20 YEAVFLIIIV-----YLFERNILFFPKRYPKNTPKIGVSN 58  
Qy 74 TIGSQLPNASINVENFLIITAGPLIYRVFAPVCENIVNLTILIVILEPTDSMSWFF 133  
Db 59 TTAMIIAVAVSVVVLMLAGGLTAALFRGYPGFRVLTLELIVKISGL-----LF 109  
Qy 134 WVTIGM--ATSNFNSGLYNSVGVGGDFPHTYIGALLIGNNICGLL----- 179  
Db 110 GPIIGTSAATIDFLTIVISGCVFNIG-----YVIGAILTG-MIAGILREVLISTSFLNN 163







```

D0      13  YSVLFWCLILSFSVLN--EMVLNVSLPDIANHTTGGTNVNVNTAYMLTTSISGPAVYG  70
QY      82  ASINVENL-FLIIAGPLLYRVFAPCFNVNLTILILVILEPTEDSWSFWFVTLGMA  140
D0      71  KLSDYINIKKLLIIG-----ISUSCLGSLTAFI-----GNHFFILIFG--  109
QY      141  TSINFSNGLYENSIVGYG-GDFEHT-----YIGALL-----IGNNI  175

```

	Qy	176	CGL-----LITVVGKGYTYFINDPKIVAIVYFGISL-----VILLVC	213
		:	:   :	:
	Db	159	GGIHAIHWSSLLLPMTIIVI-----PFLIKVMVPGSKNTLDIVGVLMSI	209
		:	:   :	:
	Qy	214	AIALFFITTKODFYHHQKGMEIREKAETDRPSILWTTFTNCYGLENVFPCPAVTLT	273
		:	:   :	:
	Db	210	SICEML-----FTTNWNTELIIDPTI-----	231
		:	:   :	:
	Qy	274	IFPVMTITVTRGDSGFLNKIMSEN-DEITYTLTSLFVNLFALIGSVASKIHWPTRYL	332
		:	:   :	:
	Db	232	FVFIKHISRVSNPFINPKLGRIIPMCLGSGLIFSIVAGFSMVPMYM-KTIYHV	289
		:	:   :	:
	Qy	333	KFAIIIRALPIP-----PPFCNRYRVOTRAYVPVFESTDIPVIGIAMSFHGYSLSA	385
		:	:   :	:
	Db	290	NVATIGNSVLFPGTMSVIYVGGELVDKRG-----SLEFVILGLSIS-----ISFL	338
		:	:   :	:

```

Query Match      4.9%; Score 113; DB 1; Length 614;
Best Local Similarity 19.3%; Pred. No. 0.65;
Matches 98; Conservative 66; Mismatches 149; Indels 196; Gaps
QY 21 KYNIVYVWLVLVGVGLLPWNMEITIA--PEYVYV----- 54
||| :||| :||| :||| :
16 KYNFEFFVILIL-----FTITLATPPFVGRGNHAGNDAFNKVARVWSTISALKDQ 64

```

```

QY 55 ----WFKPDGVE-----TWYSKEPFMGSLTIGSQLPNASINVENFLIIAGPLIYRVFAPVC 106
Db 65 QVTPQDPDNALSGFGYAW--NEEVG-----PLPTYFISVIK-FIVKSWLSLSFLYSLC 115
QY 107 FNVNLTIIILIVILEPTEDSMWPFVTLGMATSNFNGLYENSIVYGVGGDFPHYI 166
Db 116 LFTSGIFIFNFSSFLKDHNTS-KLGLLAVALT---FSNSTYINLYYYANPSQPLALL 171
QY 167 GALLI--GNN-----ICG-----LLITVVKIGVT 188
Db 172 FVILLFWMKMKNKRSFAFLMVAAGAAGLPLSHVTTCITLFPVLLYLLFLIIRKG-- 229
QY 189 YFLNDEKLVAIYVFGISLVILLVCAIALEF-----ITKQDFYH----- 227
Db 230 -NKENIKIIGLGLSVTSAI---GLSAFFLPLENLAKSGIYVNSNSDFSRSGWNNI 284
QY 228 -YHQKGMETREKAETDRSPSILWTF-----TNC-YGQLFNWVFCFAVT 271
Db 285 AYFGKWEPLYKIEFSYKFPSSLFVILFIFISLINEKKTNAKYSLIFS---CFSLV 341
QY 272 LTIPFVMMVTTRGDSGLNKMSENDEIYTLTTS---FLVENLFAAIG-----SIVAS 322
Db 342 LVLMLPI-----FPWKLES---IFTIVQDPARFSTLGLFSALSLLVILPILLD 388
QY 323 KIHWPTRYLKFAIILRALPIPEFFFCNYRVQTRAYPVF-----FESTDIF 368
Db 389 KISGKTSYLTIGLLVIFSLGFAERN-RIQKGSQPLFASQSLNKNKTPFNMYMENPDSI 447
QY 369 VIGGIAMSFHGYLSALAMGYTPNVVPSH 397
Db 448 AIG-----EYLPQVIGSH 460

```

Search completed: February 27, 2002, 17:02:49  
Job time: 195 sec

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	2265	98.7	434	2	T15080	hypothetical prote
2	716.5	31.2	461	2	T28006	hypothetical prote
3	706.5	30.8	450	2	T23528	hypothetical prote
4	574.5	25.0	800	2	T23247	hypothetical prote
5	313.5	13.7	428	2	T21489	hypothetical prote
6	249.5	10.9	408	2	C86156	hypothetical prote
7	214	9.3	327	2	JC4196	36K hydrophobic nu
8	209.5	9.1	326	2	JC4195	36K hydrophobic nu
9	196.5	8.6	513	2	T23887	hypothetical prote
10	182.5	8.0	517	2	S36712	FUN26 protein - ye
11	180	7.8	143	2	T22164	hypothetical prote
12	144	6.3	418	2	D85064	hypothetical prote
13	138.5	6.0	418	2	E85064	hypothetical prote
14	134	5.8	432	2	H64383	Na+ transporter -
15	128.5	5.6	382	2	G96541	hypothetical prote
16	126	5.5	418	2	F83986	transporter BH2694
17	125.5	5.5	674	2	T21217	hypothetical prote
18	117.5	5.1	506	1	D64048	iron (III) ABC tra
19	117.5	5.1	559	2	S62503	inorganic phosphat
20	115	5.0	461	2	B83601	probable transport
21	114.5	5.0	449	2	T35053	DNA damage-inducib
22	113.5	4.9	477	2	H64210	eggshell protein p
23	113.5	4.9	604	2	T11867	NADH dehydrogenase
24	113	4.9	459	2	S42238	tetracycline resist
25	113	4.9	614	2	S40085	hypothetical prote
26	112.5	4.9	694	2	E69143	hypothetical prote
27	112	4.9	383	2	A42111	Na+/H+-exchanging
28	111.5	4.9	387	2	B64459	Na+/H+-exchanging
29	111.5	4.9	398	2	S66518	proteinase-activat

```
|||||
Db 361 FFESTDIFVIGGIAMSFSGYLSALAMGYTPNVVPSHYSRFAAQLSVCTLMVGLLTGGLW 420
QY 421 PVVIEHFVDKPSIL 434
|||||
Db 421 AVVIEHFVDKPSIL 434

RESULT 2
T28006
hypothetical protein ZK809.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28006
R:Dobson, R.
submitted to the EMBL Data Library, December 1995
A:Reference number: Z20454
A:Accession: T28006
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-461 <WIL>
A:Cross-references: EMBL:Z68303; PIDN:CAA92642.1; GSPDB:GN00022; CESP:ZK809.4
A:Experimental source: clone ZK809
C:Genetics:
A:Gene: CESP:ZK809.4
A:Map position: 4
A:Introns: 54/3; 110/2; 142/1; 357/3
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK809.4

Query Match 31.2%; Score 716.5; DB 2; Length 461;
Best Local Similarity 35.8%; Pred. No. 4.5e-49;
Matches 163; Conservative 86; Mismatches 167; Indels 39; Gaps 9;
QY 7 SNTYAVEQEA-FPRDKYNIYVWLIVGVGVLPPWNMFITIAPEYVYVYVWF-KPD----- 59
Db 11 NKTAKVEEPEPEKDKGNLVFIILHIGITLMPWNMLTISDYFESYKMLANSTIDMD 70
QY 60 -----GVETWYSKEPFMGSLTIGSOLPNASINVFNLIIAGPLTYRVFAPVCFNIVNLTI 114
Db 71 TGVVVTGYPTVYSSNFQSFQTIASQVPLNLLNLLNIFIVKGLASRI--TVGLSIVAVCV 128
QY 115 ILILVIVLEPTEDSMWFFWTLMGATSNFNGLYENSVYGVGGDPHPHYIGALLIGNN 174
Db 129 ITTMMFIVTSTWLTGFTTITIIIVLNGANGVYQNSIFGLASELPFKYTNNAVIIGN 188
QY 175 ICGLLIIVWKI---GVTFYFLNDEPKLVAIVYFGISLVILLVCAIALFFITKQDFYHVHHQ 231
Db 189 LCGTFVLLSMSTKAVTRNILDR-----SFAYFSIALTLVFCFISFHLKKQRFQYIST 244
QY 232 KGMBIREK---AETDRPSPSILMTTTCYQOLFNFVFCFAVTLTIPP-VMMTVTTTRGDS 287
Db 245 RAERQRNKNDKDEAVDSEKGVANYIATPKFAFPOLINVLVFFVTLSTIFPGVMYVVKDEKG 304
QY 288 G-----FLNKIMSENDEI-----YTLTSLFVLNFAAGSIVASIKHWPTRYLK 333
Db 305 GYVDPLPQROFFNTSLIHNREFSEKTFMDVTTFLQNFVFAFGSIVAGKQWPAKPLW 364
QY 334 FAITLRALFIPFFFCNRYRQTRAYPVFFESTDIFVIGGIAMSFSGYLSALAMGYTPNV 393
Db 365 IPVYLRLLYTPFFIFCNLPETSLPVPFFESTWLVFIIAASMFSGYFSGLAMWYTSKT 424
QY 394 VPSHYSRFAAQLSVCTLMVGLLTGGLMPVVIHFV 428
Db 425 VDPKSAQVAGMMAGFFLISIGVSLFTMTVMKV 459

RESULT 3
T23528
hypothetical protein K09A9.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T23528
```

```
R:Swirnburne, J.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19753
A:Accession: T23528
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-450 <WIL>
A:Cross-references: EMBL:Z79601; PIDN:CAB01882.1; GSPDB:GN00028; CESP:K09A9.3
A:Experimental source: clone K09A9
C:Genetics:
A:Gene: CESP:K09A9.3
A:Map position: X
A:Introns: 59/3; 115/2; 147/1; 318/1; 346/3
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK809.4

Query Match 30.8%; Score 706.5; DB 2; Length 450;
Best Local Similarity 36.0%; Pred. No. 2.7e-48;
Matches 157; Conservative 88; Mismatches 164; Indels 27; Gaps 8;
QY 12 VEQEA-FPRDKYNIYVWLIVGVGVLPPWNMFITIAPEYVYVYVWF-----KPD 59
Db 21 VEETPEPEKDKGNLVFIILHIGITLMPWNMLTISDYFESYKMLANSTIDMDTGKVT 80
QY 60 GVETWYSKEPFMGSLTIGSOLPNASINVFNLIIAGPLTYRVFAPVCFNIVNLTIILV 119
Db 81 GDPTVYSSNFQSFQTIASQVPLNLLNLLNIFIVKGLAGRI--TVGLSIVAVCVITMI 138
QY 120 IVLEPTEDSMWFFWTLMGATSNFNGLYENSVYGVGGDPHPHYIGALLIGNICGLL 179
Db 139 FIVETSTWLTGFTTITIIIVLNGANGVYQNSIFGLASELPFKYTNNAVIIGNLCGTF 198
QY 180 ITVVKI---GVTFYFLNDEPKLVAIVYFGISLVILLVCAIALFFITKQDFYHVHHQKMEI 236
Db 199 VTLLSMSTKAVTRNILDR-----SFAYFSIALTLVFCFISFHLKKQRFQYSTRERQ 254
QY 237 R---EKAETDRPSPSILMTTTCYQOLFNFVFCFAVTLTIPP-VMMTVTTTRGDSGFLNK 292
Db 255 RAKNEEAADNEGRMANYIATPKFAFPOLINVLVFFVTLSTIFPGVMYVVKDEKGGTYDF 314
QY 293 IMSENDEIYTLTSLFVLNFAAGSIVASIKHWPTRYLKFAILLRALFIPFFFCNRYR 352
Db 315 PLPON--YFMDVTTFLQNFVFAFGSIVAGKQWPAKPLWIPVYLRLLYTPFFIFCNLY 372
QY 353 VQTRAYPVFFFCNRYRQTRAYPVFFESTDIFVIGGIAMSFSGYLSALAMGYTPNVVPSHY 412
Db 373 PETRLPVFFESTWLVFIIAASMFSGYFSGLSMAYTSKTVDPKSAQVAGMMAGPFLIS 432
QY 413 GLITGGLMPVVIHFV 428
Db 433 GIVSGLIFTMTVIRFV 448

RESULT 4
T23247
hypothetical protein K02E11.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23247
R:McMurray, A.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z19715
A:Accession: T23247
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-800 <WIL>
A:Cross-references: EMBL:Z77665; PIDN:CAB01223.1; GSPDB:GN00023; CESP:K02E11.1
A:Experimental source: clone K02E11
C:Genetics:
A:Gene: CESP:K02E11.1
A:Map position: 5
A:Introns: 63/3; 80/3; 120/3; 137/1; 178/1; 280/3; 333/3; 418/3; 464/2; 496/1; 521/2;
```

Query Match 25.0%; Score 574.5; DB 2; Length 800;  
Best Local Similarity 32.7%; Pred. No. 1.3e-37;  
Matches 134; Conservative 88; Mismatches 165; Indels 23; Gaps 10;

Qy 19 RDKYNIVVWLVLGVGVLPPNNMFTTIAPEYVYVWPK-PDGVETWYSKEFMGSLTIGS 77  
Db 388 KDLGNVVFIFPMFMFGALLPNNMFLNLSFDYTMFKLRSADGNATWYSSNFQSMTISA 447  
Qy 78 QLPNASINVENFLIIAGPLIYRFAPVCFNIVNLTIILIVLIVLEPTEDSMSEFFWTL 137  
Db 448 QIPSLVFSVINIFAVKGLDTRGM--KCLIVVQVMVTVTVYIIDTWTIAFTSMULT 505  
Qy 138 GMATISINFSNGLYSVYGVGDPEHTYIGALLIGNNICGLLTIIVKIGVYFVFLNDEPK 197  
Db 506 GTIVVLNAANGLFQNSMGLSPPEPKYTNNAVIGQNFCCGTAVTVLSM-LTKAASDDVQM 564  
Qy 198 VAIVFGISLVILLVCATALFTFKQDYHYHHQKGMREKAEKAEKAEKAEKAEKAEKAE 253  
Db 565 RASLFFGLSSVAVVVCFLNFKRLAFY----KFGILRTSSQSDERGIS-SWESVKLA 619  
Qy 254 FTNCGQLFNWFCFAVLTITFPVMMVTYTRGDSGLFNKIMSENDEIYTLTSLFVFNLF 313  
Db 620 FEKSKMQANIFVFFVTLALFPNVCMTVKDAKKGELHSFVVP-EKYMVDVVTFVFNLF 678  
Qy 314 AAGISIVASKIHWPTRYKFAILRALFIPFFFCNRYVQ--TRAYPVFFESTDIFVIG 371  
Db 679 AFLGSLMANWIRFGPNTVVICVAARFWMFYFPAANYHPMDFPRAYPVLFESTWLFAPN 738  
Qy 372 GIAMSFSHGYSALAMGVTNPNVPSH-----YSRFAAQLSVCTLMVGLLTG 417  
Db 739 ICIFALTSGYLSLIMTYAPR---SHEDPKIQRMAGMIASFELIFGIVAG 785

## RESULT 5

hypothetical protein F1707.13 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 22-Oct-1999  
R:Vysotskaya, V.S.; Schwartz, J.R.; Toriumi, M.; Yu, G.; Kwan, A.; Oji, O.; Liu, S.; Li, rtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel, N submitted to the EMBL Data Library, June 1998  
A:Description: Arabidopsis thaliana chromosome 1 BAC F1707 sequence.  
A:Reference number: 214334  
A:Accession: T01489  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-428 <VIS>  
A:Cross-references: EMBL:AC003671; NID:g2833627; PID:g3176684; GSPDB:GN00059; ATSP:F1707  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Gene: ATSP:F1707.13  
A:Map position: 1  
A:Introns: 185/1

Query Match 13.7%; Score 313.5; DB 2; Length 428;  
Best Local Similarity 23.2%; Pred. No. 2.4e-17;  
Matches 100; Conservative 84; Mismatches 186; Indels 61; Gaps 11;

Qy 18 PRDKYNIYVWLVLGVGVLPPNNMFTTIAPEYVYVWPKPDGVETWYSKEFMGSLTIGS 77  
Db 35 PSDSYHFAIIVFTLGVGFLPPNNAFITAVID--YFSYLPSTAVDRIFAIVYM----- 85  
Qy 78 QLPNASINVENFLIIA-----GPLIYRFAPVCNIVNLTIILIVLIVLEPTEDSMSEWF 132  
Db 86 -----LVALVCLFVIVVFYAHKSLASFRLNGLLLEFVALLVPPVLDLVVKGQVGLYAG 140  
Qy 133 FWTVLGMATISINFSNGLYSVYGVGDPEHTYIGALLIGNNICGLLTIIVKIGVYFVFN 192  
Db 141 FDTVSAVALSGLDALMOGGLIGVAGEMPERYMOAVVAGTAGSGVLVSLRLTKAVYP 200

Qy 193 DEP---KLVAIVYFGISLVILLVCATALFFITTKQDYHYHHQKGM-----IREKAE-----T 242  
Db 201 QDPGLRKSANLYFAVGIVVWVCAVFNVAHKLVKIFHEERKNEELIRESEKGSILT 260  
Qy 243 DRPSIILWTFNTCYGOLFNNWFCFAVTLTIFPVMVTYTRGDSGLFNKIMSENDEIYT 302  
Db 261 GLAWRTTLWDIVTKVKSFGIVLLYMTLSIFPGYIT-----EDVHS 303  
Qy 303 -LLTSF-----LVNLFAAIGSVASKIHWPTRYKFAILRALFIPFFFCNRYVQT 355  
Db 304 ELLTDWYFILLIAAYNVFDLVGKCLTAVFMLEDEKIAVGSGSTARLLFPLFGCLHG--- 360  
Qy 356 RAYPVFFESTDIFVIGGIAMSFSGHLSALAMGYTPNPNVPSHYSRFAAQLSVCTLMVGLL 415  
Db 361 ---PMFLRTEIPVTLTCLLGLTNGYLTSLVLMILAKPSVPLRHSSETAGITVTFVFLVVGLA 417  
Qy 416 TGGL--WPVVI 424  
Db 418 SGSVIAWFWVI 428

## RESULT 6

C86156  
hypothetical protein AAG10625.1 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: C86156  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon ansen, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, Chin, C.W.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: C86156  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-408 <STO>  
A:Cross-references: GB:AE005172; NID:g9972375; PIDN:AAG10625.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 10.9%; Score 249.5; DB 2; Length 408;  
Best Local Similarity 22.0%; Pred. No. 2.5e-12;  
Matches 99; Conservative 72; Mismatches 187; Indels 93; Gaps 14;

Qy 12 VEOEAFPRDKYNIYVWLVLGVGVLPPNNMFTTIAPEYVYVWPKPDGVETWYSKEFMG 71  
Db 7 IVDEVETRDAYRVAIVIHFLGAGSLIPWNAITAVD--YFGYLPDKHVEFTVYAYM- 63  
Qy 72 SLTIGSQLPNASINVENFLIIAGPLIYRFAPVCFNIVNLTIILIVLIVLEPTEDSMW 131  
Db 64 -----SCSVLVLMVMTWNTSRMSYVRMNLGFS-----MFIAMMSPLID--W 105  
Qy 132 F-----FWVLGMATISINFSNGLYSVYGVGDPEHTYIGALLIGNN----- 174  
Db 106 VMKGEKGENSVYMLMVGSVVLGGLADGVVGGSLGAGKLPRQYMQAIFAGTASSGKPSF 165  
Qy 175 -----ICGLLITVVKIGVYFVFNDEP---KLVAIVYFGISLVILLVCATALFF 219  
Db 166 LCKTFLVLLSFLPGLIIISLLRIATKASLPQTPQCMRTSAHSYFIVSTLILCCFICNV 225  
Qy 220 ITKQDYHYH---HOKGMEIREKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 274  
Db 226 LHKLPVMOQHLKFHQ-----PLHSTLTIWVGRKIKWPASCMLLIYSVTLSI 272  
Qy 275 FPMVTYTRGDSGLF-NKIMSENDEIYTLTSLFVFNLFAAIGSIVASKIHWPTRYLK 333  
Db 11

Db 273 FP-----GFAIENLKSQLQSWYPILLITVYNIISDFVGKSLTALYLMQSIKAT 321  
QY 334 FAITLRALFIPFFPCNRYQTRAYPVFFESTDIFVIGGIAMSFSGHYLSALAMGYTPNV 393  
Db 322 WACIVRLFLPLFSAC-----LRGPKWLTEVPVVLTFMGLTNGLYLSVLMIMAPKT 375  
QY 394 VPSHYSRFAAQLSVCTLMVGLLTGG-----LW 420  
Db 376 VHASEALAAIFMVVFLGLGLVCGSVIGLW 406

RESULT 7  
JC4196  
36K hydrophobic nucleolar protein - human  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1995 #sequence\_revision 27-Oct-1995 #text\_change 21-Jul-2000  
C:Accession: JC4196  
R:Williams, J.B.; Lanahan, A.A.  
Biochem. Biophys. Res. Commun. 213, 325-333, 1995  
A:Title: A mammalian delayed-early response gene encodes HNP36, a novel, conserved nucleolar protein  
A:Reference number: JC4195; MUID:95367016  
A:Accession: JC4196  
A:Molecule type: mRNA  
A:Residues: 1-326 <MIL>  
A:Cross-references: EMBL:X86681; NID:g951266; PIDN:CAA60380.1; PID:g951267  
A:Experimental source: heart  
C:Comment: This protein has a role in the growth response and participates in the proliferation  
C:Genetics:  
A:Gene: der12  
A:Start codon: AUG  
C:Keywords: nucleoprotein; phosphoprotein  
F:54/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predicted  
F:107/Binding site: phosphate (Thr) (covalent) (by casein kinase II) #status predicted

Query Match 9.3%; Score 214; DB 2; Length 326;  
Best Local Similarity 22.0%; Pred. No. 1.2e-09;  
Matches 71; Conservative 52; Mismatches 139; Indels 60; Gaps 8;

QY 143 INFSGLYNSVYGVGGDFPHTYIGALLIGNNICGLLITVVKI-----GVTFYLNDEPKL 197  
Db 7 INFSAVLQSLGSLFQGLGTMTSTYTLFSLGQGLAGIFAAALMLLSMASGV-----DAET 60

QY 198 VAIVYF---GISVLVLCVLAIAFFITKODFY---HYHKGKMEIREKA----- 240  
Db 61 SALGYFTTPVVGILMSVLCYLSPHLKFARYLANKSSQAQAELETKAELLQSDENGIP 120

QY 241 -----ETDRPSPSILWTFNCTYCGQLFNWVFCFVTLTIFFPV 277  
Db 121 SSPQKVALTLDLQLEKEPESEPEQPKGKPSVTFVQKIWLTAALCLVLTFTVLSVFP 180

QY 278 MMTVTTRGDSGLNKMSENDEIYTLTSLFVNLFAAIGSIVASKIHP--TPRYLKFA 335  
Db 181 ITAMVTSSTS-----PGKWSQFNPCICFLNIMDLGRSITSYFLWDEDSRLPL 234

QY 336 IILRALFIPFFPCNRYQTRAYPVFFESTDIFVIGGIAMSFSGHYLSALAMGYTPNVVP 395  
Db 235 VCLRFELFVPLFMLCHVPQSRK-LPILFPQDAYFITFMLLFAVNSGYLSVLTMTCLAPQVL 293

QY 396 SHYSRFAAQLSVCTLMVGLLTG 417  
Db 294 PHEREVAGALMTFFLAGLSCG 315

RESULT 8  
JC4195  
36K hydrophobic nucleolar protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Sep-1995 #sequence\_revision 27-Oct-1995 #text\_change 05-Nov-1999  
C:Accession: JC4195  
R:Williams, J.B.; Lanahan, A.A.  
Biochem. Biophys. Res. Commun. 213, 325-333, 1995  
A:Title: A mammalian delayed-early response gene encodes HNP36, a novel, conserved nucleolar protein

A:Reference number: JC4195; MUID:95367016  
A:Accession: JC4195  
A:Molecule type: mRNA  
A:Residues: 1-327 <MIL>  
A:Cross-references: EMBL:X86682; NID:g951302; PIDN:CAA60381.1; PID:g951303  
C:Comment: This protein has a role in the growth response and participates in the proliferation  
C:Genetics:  
A:Gene: der12  
A:Start codon: AUG  
C:Keywords: nucleoprotein; phosphoprotein  
F:52-159/Domain: hydrophilic #status predicted <HYD>  
F:54/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predicted  
F:107/Binding site: phosphate (Thr) (covalent) (by casein kinase II) #status predicted

Query Match 9.1%; Score 209.5; DB 2; Length 327;  
Best Local Similarity 22.4%; Pred. No. 2.8e-09;  
Matches 76; Conservative 49; Mismatches 134; Indels 81; Gaps 12;

QY 131 WFEVWTLGMATSFNFSNGLYNSVYGVGGDFPHTYIGALLIGNNICGLLITVVKI----- 185  
Db 5 WF-----INSECAVLQSLGSLFQGLGTMTSTYTLFSLGQGLAGIFAAALMLSLAS 54

QY 186 GVTFYLNDEPKLVAIVYF---GISVLVLCVLAIAAL-----PFI-----TK 222  
Db 55 GV-----DAQTSALGYFITPCVIGILLSIVCYLSLPHLKFARYVLTETKLSQAPTQLETK 108

QY 223 QDFYHYHKGKMEIR-----EKAETDRPSPSILWTFNCTYCGQLFNWVF 266  
Db 109 AELLQADEKNGVPLSPQQAASPTLDLQLEKEPEEPEQPKGKPSVVFVFRIMLTALCLVL 168

QY 267 CFAVTLTIFFVM---MTVTTRGDSG----FLNKMSENDEIYTLTSLFVNLFAAIGSIV 320  
Db 169 VFTVLSVFPFAITAMVTTSSNSPGKGLFFNPI-----CCFLFNVDMLGRSL 217

QY 321 ASKIHP---TPRYLKFAIILRALFIPFFPCNRYQTRAYPVFFESTDIFVIGGIAMSF 377  
Db 218 TSFYLPMPDEDSQQLPLLVCLREFVPLFVPLMCHVPQHAR-LPIIFRODAYFITFMLLFAV 276

QY 378 SHGYLSALAMGYTPNVVPSHYSRFAAQLSVCTLMVGLLTG 417  
Db 277 SNGYLVSLTMC LAPRQVLPHEREVAGALMTFFLAGLSCG 316

RESULT 9  
T21887  
hypothetical protein F36H2.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T21887  
R:Steward, C.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19483  
A:Accession: T21887  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-513 <MIL>  
A:Cross-references: EMBL:Z81078; PIDN:CA03075.1; GSPDB:GN00019; CESP:F36H2.2  
A:Experimental source: clone F36H2  
C:Genetics:  
A:Gene: CESP:F36H2.2  
A:Map position: 1  
A:Introns: 28/3; 56/1; 93/2; 163/3; 277/3; 314/1; 341/3; 384/2; 407/2; 458/2; 480/3

Query Match 8.6%; Score 196.5; DB 2; Length 513;  
Best Local Similarity 21.4%; Pred. No. 4.9e-08;  
Matches 95; Conservative 77; Mismatches 159; Indels 113; Gaps 21;

QY 31 LVGFGVLLPWNMFITTAPEYVYVNWFEKPGCV-----ETWYKEFGSLTIGSQLP 80  
Db 11 IVSISMFNPNAFNN-AHEY---FYKLRNVTNDEDPVNTDTTWFIKR-----QIRDDVP 61

Qy 81 NASINVENLFLIAGPL-----IYRVFAPVCNIVNLTII-----LILVIVLEPT----- 125  
Db 62 TELQANFEAYLTVYGSIACVLGSILNVPATK--SLSNSTRMIGWGHILVVVFIPITALT 119  
Qy 126 ---EDSMWFVWTLGMAITSINF-SNGLYENSVYGVGGDFPHYTGALLIGNICGLLIT 181  
Db 120 WNFDDQDQFFFNLSMILSIACFGSLGMLAGVGLSALFPSSQYTOAVMVGQSFAGVLA 179  
Qy 182 VVKI---GVT---YFUNDPEKLVAVYFGISLVILLVCAIALFFIT-----KODF 225  
Db 180 LMSILCOAVTSDVILNGQ-----MYFGSLIMCFISLTIATYLLTPPMITDDGSGEL 233  
Qy 226 YHYHOKGMEIR-----EKAETDRPSPSILWTFITNCYGOLENVWFCAVTL 272  
Db 234 IENEVESEIAQANHFPPIDSDNSQTEEHQLPK---WTMYTDIIRK-----SAIDL 282  
Qy 273 TIFPVMVTVTRGDSGFLNKINSEN-----DEIYTLLSFLVFNFAAGISVASKIHP 327  
Db 283 TTISVVLIVTLAAYPGLTSLVHSTSRNHTWNSYFSAVASFLLYNVGDLIGRSSANSRL- 341  
Qy 328 TPRLKFAILIRALFIPFFFCN-----YRVQTRAYVPVFESTD-----IFVIG 371  
Db 342 SPKYLLIISFLRFALIPMIAMCNVHILMHLFMTVLEFLSIDSIDKDSRELAGSIALF 401  
Qy 372 GTAMS-----FSHGYSALAMGYT 390  
Db 402 GVTAIFISFFSICHFSCISIVYT 425  
RESULT 10  
FUN26 protein - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YAL022c  
C:Species: Saccharomyces cerevisiae  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 29-Oct-1999  
R:Oucllette, F.; Clark, M.W.; Keng, T.; Storms, R.K.; Zhong, W.; Zeng, B.; Fortin, N.; D  
submitted to the EMBL Data Library, January 1993  
A:Description: Sequencing of Chromosome I from Saccharomyces cerevisiae: analysis of a 3  
A:Reference number: S36711  
A:Accession: S36712  
A:Molecule type: DNA  
A:Residues: 1-517 <OUE>  
A:Cross-references: EMBL:L05146; NID:g171851; PIDN:AAC04935.1; PID:g171853; MIPS:YAL022c  
C:Genetics:  
A:Gene: SGD:FUN26  
A:Cross-references: SGD:S0000020; MIPS:YAL022c  
A:Map position: 1L  
C:Keywords: transmembrane protein  
F:79-95/Domain: transmembrane #status predicted <TM1>  
F:155-171/Domain: transmembrane #status predicted <TM2>  
F:177-193/Domain: transmembrane #status predicted <TM3>  
F:218-234/Domain: transmembrane #status predicted <TM4>  
F:245-261/Domain: transmembrane #status predicted <TM5>  
F:344-360/Domain: transmembrane #status predicted <TM6>  
F:412-428/Domain: transmembrane #status predicted <TM7>  
F:445-461/Domain: transmembrane #status predicted <TM8>  
F:500-516/Domain: transmembrane #status predicted <TM9>

Db 172 -PKWFNFIMLVVVISSMGTAQTQIMAIANVFGESEYSGQVMVGOAVAGVLPSSLVLF 230  
Qy 187 VTYFLNDEPKLVA--IVYFGISLVILLVCAIALFFITK----- 222  
Db 231 LAFIENSSVSTGGILLVFFFTTLLVVVIC-VVMFVSYSKISRKVRKNENMNVEDGHITDVLIG 289  
Qy 223 -----QDFVHYH-----HOKGMEIREKAETDRPSPSILWTFITNC 257  
Db 290 SLRSENEEIRIVGRIDQDEDEHRTNGTRDDNDGGELOLKVPE-----VLFAKL 341  
Qy 258 YQLENVWFCAVTLITIFPVMVTVTRGDSGFLNKIMSENDEIYTLTSFLVFNLPAAIG 317  
Db 342 KYLVLSIETFEVVTL-VFVPFASAT-----YVTGLPSNAQYIPLI--FTLWNLGDLYG 392  
Qy 318 SIVASKIHP-----TPRYLKFAILIRALFIPFF 346  
Db 393 RVIAD---WPMFRDQKFTPRKFTIYSLLRVAAIPLF 425  
RESULT 11  
T22164  
hypothetical protein F44D12.9 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T22164  
R:Coles, L.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: Z19525  
A:Accession: T22164  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-143 <WIL>  
A:Cross-references: EMBL:Z68298; PIDN:CAA92605.1; GSPDB:GN00022; CESP:F44D12.9  
A:Experimental source: clone F44D12  
C:Genetics:  
A:Gene: CESP:F44D12.9  
A:Map position: 4  
A:Introns: 19/1; 37/2  
Query Match 7.8%; Score 180; DB 2; Length 143;  
Best Local Similarity 40.0%; Pred. No. 2.4e-07;  
Matches 32; Conservative 18; Mismatches 26; Indels 4; Gaps 1;  
Qy 14 QZAFPRDKYNIYVWLVLVGVGLLPWNMFITTAPEYVYVYVWFKPDGVETWYSKEPMGSL 73  
Db 67 KESSPTDRWLYVIITFMHGMGLMSWNMFITTAPOYHYHWFN-----NTNYODSFMSII 122  
Qy 74 TIGSQLPNASINVFNLFII 93  
Db 123 GVTSQLPNVGMILNTIVVM 142  
RESULT 12  
D85064  
hypothetical protein AT4G05120 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
R:anonymouse, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp  
Nature 402, 769-777, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488  
A:Accession: D85064  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-418 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7267271; PIDN:CAB81054.1; GSPDB:GN00140  
C:Genetics:  
A:Gene: AT4G05120  
A:Map position: 4

Query Match 6.3%; Score 144; DB 2; Length 418;  
Best Local Similarity 20.6%; Pred. No. 0.00054;  
Matches 96; Conservative 76; Mismatches 159; Indels 134; Gaps 23;

QY 18 PRDKYNIVVWLVI---LVGFGVLLPWNFFITAEYY-VNYWFKPDGVETWYSKEF-MGS 72  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 9 PPEKQGKIYQAMVCCILIGISLWSNMLTIADYYKYVPDPYHPSRVLTLYVQPFALGT 68  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 73 LTI-----GSQLPNASINVFNLFIAGPLIYRFPVCFNIVNLTIIILIVILEPTEDS 128  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 69 ILILAYHESKINTRKRNLGYIL-----FTISTFLIIVLD----- 103  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 129 MSWFPPWTLGMAAT-----SINFNGUYENSVMG--VGG-----DEPHTYIG 167  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 104 -----LATKGRRGIGPYIGLCVVAVSGFLADATVOGGMICDLSLMCPELVSQFMG 153  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 168 ALLIGNNICGLLITVKVIGTVFYFNDEPKVAIVYFGISVLILLVCAIALFEI-TKODFY 226  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 154 GLAVSGAITSALRLITKAAFEK-TNDGRKKAMFELAISTCIELLCVFLYAIVFPKLPIV 212  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 227 HYHQKHGMEIREK-----AETDRSPSILMTTFTNCYGOLFENVW 265  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 213 KYRRAKAASEGSKTVSADLAAGIQNSDLTDSDSKNQRLSNKELLIGNIDYAVNFLIY 272  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 266 FCFAYTLIFPMMVTTRGDSGFLENKIMSEN--DEITYTLTSFLVNFLFAIGSVASK 323  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 273 VC---TLSIFF-----GFLYENTGGHGLDWMYLV-LVAMYNCMDLVGR----- 312  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 324 THWPTRYLK-----FAILRALFIPFFFCNRYVOTRAYPVVFESTDIFVIGIAM 375  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 313 -YTPLVKKWKTKENRKLIITIAVLSRYLLIPAYF-TAKYGDGQWMIML-----ISVLG---- 362  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 376 SFSHGYSALANGYTPNVVPSHYSRFAAQLSVCTLMVGLLTGGLW 420  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 363 -LTNGLHTVTCIM---TIAPRGY-KGPEQNALGNLLVIFLLGGIF 401  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 13

E85064  
hypothetical protein AT4g05130 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: E85064  
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488  
A:Accession: E85064  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-418 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7267272; PIDN:CAB81055.1; GSPDB:GN00140  
C:Genetics:  
A:Gene: AT4g05130  
A:Map position: 4

	Query Match	6.0%;	Score 138.5;	DB 2;	Length 418;
	Best Local Similarity	21.0%;	Pred. No. 0.0015;		
	Matches	95;	Conservative	73;	Mismatches 174; Indels 111; Gaps 23;
Qy	21 KYNIYVWLVLVGFGVLLPNMFTIAPXY-VNYWFKPGCVETW-YSKBFMGSLTIGSQ	78			
Dd	16 KIQAWVCCL-GIGSUFWSNMSTIIADYYQVPDPDHPKSRVFLTIYQPIALGTIMT---	71			
Qy	79 LPNASINVFNLFLIIAGPLIVRPAFCFNVNITILILVILEPTEDSMNSFFFWTLG	138			
Dd	72 LAYRESKISTRKILTGYLTF-----TISTFLVLVDLT-----TKG	108			
Qy	139 -----MATSFNSGLYENSVY-GVGDFP---HTYICALLIGNNICGLITVVVKIG	186			
Dd	109 HGGIGHYITVLCTIIVASRGELADATVKGGVLGDLSLMCPELIQSYMAGSGMAGALTSLVRL-	167			

```

Qy 187 VTVEL-----NDEPKLVAIVREGISLVLLVCAIALFET-TKQDYHYHHOKG----- 233
   :|      |::: ||||| ::| :| :|
Db 168 ITKAFAEKSNLSRKGAMIFLAISTFIELLCVLVAYVFVKPLPIVKYRYRKAASEGSKTV 227
   :|      |::: ||||| ::| :| :|
Qy 234 -----MEIRKAETD-----RPSPSILWTFETTCYGOLENFVWFCAVTLTIPVM 278
   :|      |::: ||||| ::| :| :|
Db 228 VADLAAAGIONLSLDSDSKNQMLRKELLQN-----IDHAVNLFLIYVLTLSIEP-- 280
   :|      |::: ||||| ::| :| :|
Qy 279 MVTTRGDGSGLANKIMSEN--DEITYTL--TSFLVFNLFRAIGSIVASKIHV----PTPRP 331
   ||||| :| :| :| :| :| :|
Db 281 -----GLFYENTGQHGLDGWYAILVATYTNFWDLFGRYAPLV---KWKLNRKA 327
   :|      |::: ||||| ::| :| :|
Qy 332 LKFAILLRALPIFFFCNYRVQRAYVPVFESTDIEVIGGIAMSFSGYLVSALAMGYTP 391
   ||||| :| :| :| :| :| :|
Db 328 LTIAVLTRYFLVPAFYP-TAKYGDKGMIMLVS-----ILGLTGHLTVCIMTIAP 377
   :|      |::: ||||| ::| :| :|
Qy 392 NVVPSHYSRFQAQLSVCTLM-----VGLLTGSLW 420
   ||||| :| :| :| :| :|
Db 378 NGYKPEKNALGNLLVFIILGGAVVGISLGWLW 410
   :|      |::: ||||| ::| :| :|

RESULT 14
H64383
Na+ transporter - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 31-Dec-2000
C:Accession: H64383
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak-
rson, J.C.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc-
A:Reference number: A64300; MUID:96337999
A:Accession: H64383
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-432 <BUL>
A:Cross-references: GB:U67514; GB:L77117; NID:q2826304; PIDN:AAB98666.1; PID:g1591386
C:Genetics:
A:Map position: REV598993-597695
A:Superfamily: probable transporter MJ0672

```

```

Query Match          5.8%; Score 134; DB 2; Length 432;
Best Local Similarity 19.1%; Pred. NO. 0.0034;
Matches 82; Conservative 80; Mismatches 133; Indels 134; Gaps 22;

Qy 66 SKPEWG-----SLTIGSQLPNASINVFNFLLIAGPLIYRFAPVCFNIVNLTITILV 119
      |||||  ||  |||  :  ||||  |||  :  :  :  :  :  :  :  :  :
Db 4  SKEFIGLITASLIFGSSLPDIYKGI--VILIVAGCLWFELLPLPVTSLAHPIMAVFL 61
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy 120 IVLEPTEDSNWF----FWVTLG--MATSINFSN--GLYENSIVYGVGDFPHIT----- 164
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 62 GIFN-LKEALTYFAHPILFELGFMQAQALKNHNDKFIAYKLLNYGDKFKTCTPLMFL 120
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy 165 --YIGALLIGNICGLLITVVKIGTV-----FLENDEPKL 197
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 121 SAYFLSMWISNTSATLILLPIALGLLHKTKGLRDFLLGLVAYSASIGGIATIGSPNA 180
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy 198 VAIYV-----FGISLVILLVCAIALF-----FITKODFYHHQKMEIREKA 240
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 181 IASSYLDYGFSSFKWGGPPLISLLFLICTLTLYIFKKWIPKED-----IAIQARM 231
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy 241 ETRDPSPSI-----LWTTETNCYQOLFNVWFCFAVTLTIFFVMMVTVTRGDSGFLN 291
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 232 ELSRNAVKLLVIFVLIASLW-IISDYLSEIFNVQY-FDSVIAIFAILL-----FVF 281
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy 292 KIMSD-----EYITLL-----TSFLVNLFAATGISVASKIHMPTR 330
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 282 NLVENDFKKIDMGTLLIFGLGALCLGVIVKSGANTFPLSEKLLAILGNL-----TP 333
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

- 1 -



Search completed: February 27, 2002, 17:00:18  
Job time: 204 sec

RESULT 15

G96641

hypothetical protein T25B24.2 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: G96641

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: G96641

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-382 <SPO>

A:Cross-references: GB:AE005173; NID:g4585872; PIDN:AAD25545.1; GSPDB:GN00141

C:Genetics:

A:Gene: T25B24.2

A:Map position: 1

Query Match	5.6%	Score 128.5;	DB 2;	Length 382;
Best Local Similarity	20.0%;	Pred. No. 0.0081;		
Matches	92;	Conservative	69;	Mismatches 158; Indels 141; Gaps 22;

  

Qy	20	DKYN-----IVYWLVLVIGFVLLLPNMFTITIAPEYYVNYWFKPDGVETWYSKEFMGSL	73
	:	:::::	:
	:	:::::	:
	:	:::::	:
Db	12	EKYHPSRVLTIVQYLA-----NVFI-----TLATKE-----	39
Qy	74	TIGSOLPNASINVENFLIILIAGPIIYRFAPVCNINVMT-----LILILVIVLE	123
	:	:::::	:
	:	:::::	:
	:	:::::	:
Db	40	--AKLTRLNIEFGYSLYTAGTF-----CLIIIDLASHSGSVAYVLLCLLIVAL--	87
Qy	124	PTDSMSWFVWTILGMATSFNSNGLYENSYVGVDPPHYICALLIGNNICGLLIIVV	183
	:	:::::	:
	:	:::::	:
	:	:::::	:
Db	88	-----PGLADA--FVOGAMVGDLSFMPDPF-----IQAFMAGLGIAGALTSVL	128
Qy	184	KIGVTYFLNDP---KLVAIYFGISLVILLVCAI--ALFFTCKODFYHHQKME---	235
	:	:::::	:
	:	:::::	:
	:	:::::	:
Db	129	RLITKAIFDNPSPDLRGKALLFIGITLIELACVFLYTLVFAPLPIVKYYIRAKAGKEGAK	188
Qy	236	-----TREKETDRPSPSLIWTFTTCYCOL-----FNWFCFAVLTITIPPV	277
	:	:::::	:
	:	:::::	:
	:	:::::	:
Db	189	TVSDAALAAGLOBAEQVHQHDESKIOQLTK--KOLLRENDLGINLSLIYVVLTISFP-	245
Qy	278	MWVTTRGDSGFLENKIMSEN--DEIYPLTSLFVNLFAAIGSIVAS--KIHWPTPYRL	332
	:	:::::	:
	:	:::::	:
	:	:::::	:
Db	246	-----GFLYEINTGEHLGDWYAPVLVA--MYNGWDAISRFIPSIKPLAMESRWKI	293
Qy	333	KFAIILRALPFEPFFECNRYVQTAYPVFEESTDIFVIGGIATMSFSHGCLSALAMGYTPN	392
	:	:::::	:
	:	:::::	:
	:	:::::	:
Db	294	TVCVARLLLVAPYF-TAKYADOGNWMLFIS-----FGLSNGILTVCIFTSTAPK	343
Qy	393	VVPSHYSRFAAQLSVCTLMV-----GLLTGGLPWPVVEHF	427
Db	344	GYNCPKANALGNL-MCVPFLGGITFAGVCGLWGLWLGINDSF	382

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 16:56:28 ; Search time 14.4 Seconds  
(without alignments)  
678.224 Million cell updates/sec

Title: US-08-816-011f-63  
Perfect score: 2294  
Sequence: 1 MVLINRNTYAVSEPRD.....LTGGLWPVVEHFVDKPSIL 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued\_Patents\_AA.\*
- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*
  - 3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*
  - 4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/1aa/PT05\_COMB.pep.\*
  - 6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	343.5	15.0	456	US-09-058-389A-4	Sequence 4, Appl
2	326.5	14.2	456	US-09-058-389A-2	Sequence 2, Appl
3	214	9.3	326	US-09-058-389A-3	Sequence 3, Appl
4	181	7.9	247	US-09-058-389A-8	Sequence 8, Appl
5	111.5	4.9	398	US-08-097-938-4	Sequence 4, Appl
6	111.5	4.9	398	US-08-476-000-4	Sequence 4, Appl
7	111.5	4.9	398	US-08-472-840-4	Sequence 4, Appl
8	111.5	4.9	398	US-08-476-976-4	Sequence 4, Appl
9	111.5	4.9	398	US-08-474-410-4	Sequence 4, Appl
10	111.5	4.9	398	US-08-486-673B-4	Sequence 4, Appl
11	109.5	4.8	397	US-08-486-673B-63	Sequence 63, Appl
12	108.5	4.7	398	US-08-097-938-6	Sequence 6, Appl
13	108.5	4.7	398	US-08-476-000-6	Sequence 6, Appl
14	108.5	4.7	398	US-08-472-840-6	Sequence 6, Appl
15	108.5	4.7	398	US-08-476-976-6	Sequence 6, Appl
16	108.5	4.7	398	US-08-474-410-6	Sequence 6, Appl
17	108.5	4.7	398	US-08-486-673B-6	Sequence 6, Appl
18	106.5	4.6	395	US-08-097-938-2	Sequence 2, Appl
19	106.5	4.6	395	US-08-097-938-5	Sequence 5, Appl
20	106.5	4.6	395	US-08-476-000-2	Sequence 2, Appl
21	106.5	4.6	395	US-08-476-000-5	Sequence 5, Appl
22	106.5	4.6	395	US-08-472-840-2	Sequence 2, Appl
23	106.5	4.6	395	US-08-472-840-5	Sequence 5, Appl
24	106.5	4.6	395	US-08-476-976-2	Sequence 2, Appl
25	106.5	4.6	395	US-08-476-976-5	Sequence 5, Appl
26	106.5	4.6	395	US-08-474-410-2	Sequence 2, Appl
27	106.5	4.6	395	US-08-474-410-5	Sequence 5, Appl

28	106.5	4.6	395	4	US-08-486-673B-2	Sequence 2, Appl
29	106.5	4.6	395	4	US-08-486-673B-5	Sequence 5, Appl
30	106.5	4.6	399	1	US-08-476-000-61	Sequence 61, Appl
31	106.5	4.6	399	1	US-08-472-840-61	Sequence 61, Appl
32	106.5	4.6	399	2	US-08-476-976-61	Sequence 61, Appl
33	106.5	4.6	399	3	US-08-474-410-61	Sequence 61, Appl
34	106.5	4.6	399	4	US-08-486-673B-61	Sequence 61, Appl
35	105.5	4.6	397	1	US-08-476-000-63	Sequence 63, Appl
36	105.5	4.6	397	1	US-08-472-840-63	Sequence 63, Appl
37	105.5	4.6	397	2	US-08-476-976-63	Sequence 63, Appl
38	105.5	4.6	397	3	US-08-474-410-63	Sequence 63, Appl
39	101	4.4	420	1	US-08-466-103A-2	Sequence 2, Appl
40	96	4.2	473	1	US-08-597-236-13	Sequence 13, Appl
41	96	4.2	473	1	US-08-746-682A-13	Sequence 13, Appl
42	95.5	4.2	336	3	US-08-749-816-2	Sequence 2, Appl
43	94.5	4.1	436	6	5432081-10	Patent No. 5432081
44	94	4.1	418	4	US-09-030-267-5	Sequence 5, Appl
45	93.5	4.1	569	2	US-08-750-723A-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-09-058-389A-4  
; Sequence 4, Application US/09058389A  
; Patent No. 6130065  
; GENERAL INFORMATION:  
; APPLICANT: Belt, Judith A.  
; APPLICANT: Crawford, Charles R.  
; APPLICANT: Patel, Divyen  
; TITLE OF INVENTION: A NITROBENZYLMECAPTOPURINERIBOSIDE  
; TITLE OF INVENTION: (NBMPR)-INSENSITIVE EQUILIBRATIVE, NUCLEOSIDE TRANSPORT  
; TITLE OF INVENTION: PROTEIN, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF  
; TITLE OF INVENTION: USE  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David A. Jackson, Esq.  
; STREET: 411 Hackensack Ave, Continental Plaza, 4th  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/058,389A  
; FILING DATE: April 9, 1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 1340-1-013N  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 456 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; DESCRIPTION: hENT1  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; US-09-058-389A-4



STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/058,389A  
FILING DATE: April 9, 1998  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 1340-1-013N  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 326 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: hHNP36  
HYPOTHETICAL: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-09-058-389A-3

## Query Match 9.3%; Score 214; DB 4; Length 326;

Best Local Similarity 22.0%; Pred. NO. 1.8e-13;  
Matches 71; Conservative 52; Mismatches 139; Indels 60; Gaps 8;

QY 143 INFNSGLVNSVYGGDPHYIGALLIGNNICGLITVVKI-----GVTYFLNDEPKL 197  
DB 7 INFSAVLQSGFLQGTWMPSTYTLFSGQGLAGIFAALMLLSMASGV-----DAET 60  
QY 198 VAIVYF---GISLVLLVCAIALFFITKODFY---HYHHQKQMEIREKA-----240  
DB 61 SALGVFIPTPVYGVILMSIVCYLSPLHLKARYLANKSSQAQAQAELETRAEQLQSDENGIP 120  
QY 241 -----ETDRSPSILWTFITNCYQLFNWFCFPAVTTIPPV 277  
DB 121 SSPQKVALTLDLDEKEPESEDEPKQKPSVFTVFQKIWI,TALCLVLVFTVTLSEVFA 180  
QY 278 MMTVTTRGDSGLFNKIMSNDIYTLTSLFVNLFAAIGSVIVASKIHWPP--TPRYLKFA 335  
DB 181 ITAMTSSTS-----PGKWSOFFNPICCFLEFNIMDLGRSLTSYFLWPDDEDSRLPL 234  
QY 336 IILRALFIPIFFFCNRYVQTRAYPVFFESTDIFVIGGTAMSFSGHLSALAMGYTPNVVP 395  
DB 235 VCLRELFVPLFMCHVPQSR-LPILFPQDAYFITFMLLFAVSNGLVSLTCLAPQVYL 293  
QY 396 SHYSRFAQLSVCTLMVGLTG 417  
DB 294 PHEREVAGALMTFTFLALGLSCG 315

## RESULT 4

US-09-058-389A-8  
Sequence 8, Application US/09058389A  
Patent No. 6130065  
GENERAL INFORMATION:  
APPLICANT: Belt, Judith A.  
APPLICANT: Crawford, Charles R.  
APPLICANT: Patel, Divyen  
TITLE OF INVENTION: A NITROBENZYL MERCAPTOPURINERIBOSIDE  
(NBMPR)-INSENSITIVE, EQUILIBRATIVE, NUCLEOSIDE TRANSPORT

TITLE OF INVENTION: PROTEIN, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF  
TITLE OF INVENTION: USE  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David A. Jackson, Esq.  
STREET: 411 Hackensack Ave, Continental Plaza, 4th  
STREET: Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/058,389A  
FILING DATE: April 9, 1998  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 1340-1-013N  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 247 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-09-058-389A-8

## Query Match 7.9%; Score 181; DB 4; Length 247;

Best Local Similarity 23.6%; Pred. No. 2.5e-10;  
Matches 59; Conservative 42; Mismatches 107; Indels 42; Gaps 7;

QY 18 PRDKYNIVVWLVLVGFVLLPWNMFITIAPEY-----VNTWFKPDGVE 62  
DB 7 PRDSYHLVGISFILLGLGTLLEPWFITAIPTVQARLAGAGNSTARILSTNHTGPEDAF- 65  
QY 63 TWYSEFMGSLTIGSQLPNASINVFNLIIAGPLIYRVFAPVCNIVMLTILILVIVL 122  
DB 66 -----NFNNWVTLSQLPLLETLNLSFLYQCVETVRILG----SLLAILLFAALTAAL 116  
QY 123 EPTEDSMWFVWTLGMATSFNSNGLVNSVYGGDPHYIGALLIGNNICGLITV 182  
DB 117 VKVDMSPGFFSITWASVCFINSFSAVLQSLFGQLGTMPTSTYTLFSGQGLAGIFAAL 176  
QY 183 VKI-----GVTYFLNDEPKLVAIVYF---GISLVLLVCAIALFFITKODFY---HYHHO 231  
DB 177 AMLLSMASGV-----DAETSALGVFITPCVGLMSIVCYLSPLHLKARYLANKSSQA 230  
QY 232 KQMEIREKAE 241  
DB 231 QAQELETKAE 240

## RESULT 5

US-08-097-938-4  
Sequence 4, Application US/08097938  
Patent No. 5629174  
GENERAL INFORMATION:  
APPLICANT: SUNDELIN, JOHAN  
APPLICANT: SCARBOROUGH, ROBERT M.  
TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR AND ITS  
TITLE OF INVENTION: AGONISTS AND ANTAGONISTS  
NUMBER OF SEQUENCES: 59



```

; Sequence 4, Application US/08472840
; Patent No. 5763575
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,840
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,301
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 2803-0006.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-472-840-4

Query Match 4.9%; Score 111.5; DB 1; Length 398;
Best Local Similarity 18.5%; Pred. No. 0.0041;
Matches 75; Conservative 54; Mismatches 131; Indels 145; Gaps 18;

QY 61 VETWYS-KEFMGSLTIGSQLPNASINVFNLFLIAGPLIYRVFAPVCFNIVNLTIILV 119
DB 56 VETVSVDFESASVLTGK-----LTVFLPIVYIV-----F 87
QY 120 IYLEPTEDSMNSWFF-----WTLGMAISINFGNLYENS 153
DB 88 VVGLPSNGMALWVLFRTKKHPAVIYMANLADLLSVIFLPLKTIAYHGNNTIYGEA 147
QY 154 VYGVGGDFPHTYIGALLIGNICGLLITVVKIGTVYFLNDEP-----KLVAIVVFGISLVI 209
DB 148 LCNV-----LIG-FFYGNMYCSILFWTCLSVQRVWVIVNPMGHSRKKANTAIGSLAI 199
QY 210 -LLVCAIAL-FFITKODFYHHQKMGMEIREKAETDRPSPSILWTTFTNCTY----- 258
DB 200 WLLILLVTLPIYVVKQTF-----IPAL---NITTCVDVLPQLLV 237
QY 259 GOLFNWVFCFANTLTITFPVMMVITTRGDSGFL-----NKMSENDE-----IYT 302
DB 238 GDMFNFLSLAIGVFLPFAFLTA-----SAYVLMIRMLRSSAMDENSEKRRKRAIKLIYT 328
QY 303 LITSLFVNLFAAIGSVASKIHWPTPYLKFALILRALFIPFPCNVYRVOTRAYPVFF 362
DB 293 VLAMYLII-CF-----TPSNL-----LLVHYFLIKSQGSHVYALXI 328
QY 363 ESTDIFVIGGIAMSFSGHYLSALAMGYTPNVVPSHYSRFAAQLSV 407

; Sequence 4, Application US/08476976
; Patent No. 5874400
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,976
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/390,301
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 2803-0006.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-476-976-4

Query Match 4.9%; Score 111.5; DB 2; Length 398;
Best Local Similarity 18.5%; Pred. No. 0.0041;
Matches 75; Conservative 54; Mismatches 131; Indels 145; Gaps 18;

QY 61 VETWYS-KEFMGSLTIGSQLPNASINVFNLFLIAGPLIYRVFAPVCFNIVNLTIILV 119
DB 56 VETVSVDFESASVLTGK-----LTVFLPIVYIV-----F 87
QY 120 IYLEPTEDSMNSWFF-----WTLGMAISINFGNLYENS 153
DB 88 VVGLPSNGMALWVLFRTKKHPAVIYMANLADLLSVIFLPLKTIAYHGNNTIYGEA 147
QY 154 VYGVGGDFPHTYIGALLIGNICGLLITVVKIGTVYFLNDEP-----KLVAIVVFGISLVI 209
DB 148 LCNV-----LIG-FFYGNMYCSILFWTCLSVQRVWVIVNPMGHSRKKANTAIGSLAI 199
QY 210 -LLVCAIAL-FFITKODFYHHQKMGMEIREKAETDRPSPSILWTTFTNCTY----- 258
DB 200 WLLILLVTLPIYVVKQTF-----IPAL---NITTCVDVLPQLLV 237
QY 259 GOLFNWVFCFANTLTITFPVMMVITTRGDSGFL-----NKMSENDE-----IYT 302
DB 238 GDMFNFLSLAIGVFLPFAFLTA-----SAYVLMIRMLRSSAMDENSEKRRKRAIKLIYT 292
QY 363 ESTDIFVIGGIAMSFSGHYLSALAMGYTPNVVPSHYSRFAAQLSV 407
```

```

QY 303 LITSELVENLFAAGISIVASIKHMPTRYLKAFILRALFALTPFFFCNRYNRVQTAYPVVF 362
   : | : |      || |      || |      || |      || |      || |      || :| :|
DB 293 VLAMYLI--CF-----TPSNL-----LLVVHFELIKSQSHVYALI 328

QY 363 ESTDIFVIIGGIAMSFHGYSLSALANGYTPTNVVPSPHSRFAAOLSV 407
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 329 VALCLSTUNSCIDPFVIFYFVSDFRDHAKNALLCRSVRTVKOMQV 373

RESULT          9
US-08-474-410-4
; Sequence 4, Application US/08474410
; Patent No. 6043212
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,410
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/390,301
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 2803-0006.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-474-410-4

Query Match             4.9%; Score 111.5; DB 3; Length 398;
Best Local Similarity 18.5%; Pred. No. 0.0041;
Matches       75; Conservative 54; Mismatches 131; Indels 145; Gaps 18;

QY 61 VETWYS-KEPMGSLTIGSQLPNASINVFNFLIAGPLIVRVAPVCNFVNLTIIILV 119
   || | | || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 56 VETFVSVDIFSASVLTK-----LTTVELPIVVTIV-----F 87

QY 120 IVLEPTEDSMSWF-----WVTLGMATSIKNSGLYENS 153
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 88 VVGLPNGMALVPELRTRKKKPAAIYMANLADLLSVIWFPLKIAYHIHGNNTIYGEEA 147

QY 154 VYGVGSDPHFYTGALLIGNNICLLITTVKVGYTFYNDEP-----KLVAIVFGESLVI 209
   : | : | || | | | | : | : | : | : | : | : | : | : | : | : | : |
DB 148 LCNV-----LIC-FYYGNMYCSILFTMTCLSVORYAVVINPMGHRSKRKANIAIGSLAI 199

QY 210 -LLVCAIAL-FFTIKDQFIHYHHKGMEIREKAETDRPSLSILWTFTTNCY----- 258

```



RESULT 11  
US-08-486-673B-63  
; Sequence 63, Application US/08486673B  
; Patent No. 6297026  
; GENERAL INFORMATION:  
; APPLICANT: Sundelin, Johan  
; APPLICANT: SCARBOROUGH, ROBERT M.  
; TITLE OF INVENTION: Nucleic Acids Encoding the C140 Receptor  
; FILE REFERENCE: 44481-5006-08-US  
; CURRENT APPLICATION NUMBER: US/08/486,673B  
; CURRENT FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: US 08/097,938  
; PRIOR FILING DATE: 1993-07-26  
; PRIOR APPLICATION NUMBER: PCT/US94/08536  
; PRIOR FILING DATE: 1994-07-26  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 63  
; LENGTH: 397  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-08-486-673B-63

Query Match 4.8%; Score 109.5; DB 4; Length 397;  
Best Local Similarity 19.0%; Pred. No. 0.0065;  
Matches 76; Conservative 57; Mismatches 131; Indels 137; Gaps 19;  
QY 61 VETWYS-REFMGLSLIGSOLPNASINVENFLIAGPLIYR-VFA----- 103  
DB 55 VETVFSVDEFSASVLG-----KLTTFVL-----PIVYTIIVFAVGLPSNGMALWFLF 102  
QY 104 -----PVCFNIVNTIILVILEPTEDSNWFFVTGLMATSNFNGLYENSVYGV 157  
DB 103 RTKKHHPAVIYMANALADLLSVI-----WFLPKIAYHHGNWYIGALCNV 150  
QY 158 GGDFFHTYIGALLNIGCLLITVVKIGVYFLNDEP-----KLVAIVYFGISLVILVLC 213  
DB 151 -----LIG-FYGNMYSILEMTCLSVQRYVWVNPNGHSRKRKRAIGISLAILLT 202  
QY 214 ATAL--FFITKDFHYHQQGMEIREKAETDRPSPILWTTFTNCY-----GQLF 262  
DB 203 LNVITPLVYVVKOTIF-----IPAL---NITTCDDVLPQQLVGDGF 240  
QY 263 NVWFCFAVTLTIFPVMMVTTRGDSGFL-----NKIMSENDE-----IVTLT 306  
DB 241 NYFLSLAIGVFLPAPFLTA-----SAYVLMIRLSSAMDENSEKKRRAIKLIVTVLCM 295  
QY 307 FLVFNLFRAAGSIVASKIHWPTPRYLKFAILRALFIPFFFCNRYVQTRAYVPFFESTD 366  
DB 296 YLI--CF-----TPSNL-----LLVVHYFLIKSQGSHVYALVIALC 331  
QY 367 IFVIGIAMSFSHGYSALAMGYTPNVPSHYSRFAAQLSV 407  
DB 332 LSTLNSCIDPFIYFVSHDFRDHAKNALLCRSVRTVKMQV 372

RESULT 12  
US-08-097-938-6  
; Sequence 6, Application US/08097938  
; Patent No. 5629174  
; GENERAL INFORMATION:  
; APPLICANT: SUNDELIN, JOHAN  
; APPLICANT: SCARBOROUGH, ROBERT M.  
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR AND ITS  
; AGONISTS AND ANTAGONISTS  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20006-1812

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/097,938  
; FILING DATE: 26-JUL-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 22803-20006.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 398 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-097-938-6

Query Match 4.7%; Score 108.5; DB 1; Length 398;  
Best Local Similarity 19.2%; Pred. No. 0.0083;  
Matches 75; Conservative 59; Mismatches 141; Indels 115; Gaps 18;  
QY 61 VETWYS-REFMGLSLIGSQ-----LPNASINVENFLIAGPLIYRVFA-----PVCFN 108  
DB 56 VEIVFSVDEFSASVLTGKLTTFVFLPIVYIIVPVGLPSNGMALW-VFLFRTKKHHPAVIY 114  
QY 109 IVNLTIIILVILEPTEDSNWFFVTGLMATSNFNGLYENSVYGVGGDFPHYIGA 168  
DB 115 MANLADLLSVI-----WFLPKIAYHHGNWYIGALCNV-----LIG- 154  
QY 169 LLIGNIGCLLITVVKIGVYFLNDEP---KLVAIVYFGISLVI-LLVCATAL-FFITK 222  
DB 155 FFGNMYCSILEMTCLSVQRYVWVNPNGHSRKRKRAIGISLAILLTIVTPLYVK 214  
QY 223 QDFYHYHQQGMEIREKAETDRPSPILWTTFTNCY-----GQLFNWFCFAVTLT 273  
DB 215 QTFI-----IPAL---NITTCDDVLPQQLVGDGFENYFLSLAIGVF 252  
QY 274 IFPVMMVTTRGDSGFL-----NKIMSENDE-----IYTLTSLVFNLFALIG 317  
DB 253 LFPAPFLTA-----SAYVLMIRLSSAMDENSEKKRRAIKLIVTVLAMYLI----- 299  
QY 318 STVASKIHWPTPRYLKFAILRALFIPFFFCNRYVQTRAYVPFFESTDIFVIGIAMS 377  
DB 300 CFIPSNL-----LLVVHYFLIKSQGSHVYALVIALCLSTLNSCIDPF 343  
QY 378 SHGYSALAMGYTPNVPSHYSRFAAQLSV 407  
DB 344 VYFVSHDFRDHAKNALLCRSVRTVKMQV 373

RESULT 13  
US-08-476-000-6  
; Sequence 6, Application US/08476000  
; Patent No. 5716789  
; GENERAL INFORMATION:  
; APPLICANT: SUNDELIN, JOHAN  
; APPLICANT: SCARBOROUGH, ROBERT M.  
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS  
; AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington



Sequence 6, Application US/08476976  
Patent No. 5874400  
GENERAL INFORMATION:  
APPLICANT: SUNDELIN, JOHAN  
APPLICANT: SCARBOROUGH, ROBERT M.  
TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS  
AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR  
NUMBER OF SEQUENCES: 63  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/476,976  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/390,301  
FILING DATE: 25-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: ADLER, REID G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: 2803-0006.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 398 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-476-976-6

Query Match 4.7%; Score 108.5; DB 2; Length 398;  
Best Local Similarity 19.2%; Pred. No. 0.0083;  
Matches 75; Conservative 59; Mismatches 141; Indels 115; Gaps 18;

QY 61 VETWYS-KEFGSLTIGSQ-----LPNASINVFNLIIAGPLIYRVFA-----PVCEN 108  
DB 56 VEIVSFVDEFSASVLTGKLTTPVLPVIVIVFVGLPSNGMALW-VFLERTKKKHPAVIY 114  
QY 109 IVNLTILIVIVLEPTEDSMFVWVTIGMATSNFSGLYENSIVYGVGDPPHYIGA 168  
DB 115 MANLADLLSVI-----WFLKIAVTHGNWIIYGEALCNV-----LIG- 154  
QY 169 LLIGNNICGLITVVKIGTYFINDEP----KLVAIVYFGISLVI--LLVCAIAL-FFITK 222  
DB 155 FFGNNMYCSILEFTCLSVQRYVIVNPMGHSRKKANIAIGISLAIWLLILLTYPIYVVK 214  
QY 223 QDPYHYHOKGMEIRKAEKTDPRSPSILWTFITNCF-----GOLFNVWFCFAVTLT 273  
DB 215 QTIF-----IPAL-----NITTCVDLPEQVLVGMENYFSLAIGVF 252  
QY 274 IFPVMVTVTTRGDSGFL-----NKIMSENDE-----IYTLTSTFLVNLFAAIG 317  
DB 253 LFPAPLTA-----SAYVLMIRMLRSSAMDENSEKKRAIKLIVTVLAMYLI----- 299  
QY 318 SIIVAKIHWPTPRYLKFAILIRALFIPFFFCNYRVOTRAYPVFEFSTDFIVIGIAMSF 377  
DB 300 CFIPSNL-----LLVVHYFLIKSQCSQSHVYALYIVALCLSTLNSCIDPF 343  
QY 378 SHGYLSALAMGYTPNVVPSHYSRFAAQLSV 407

Db 344 VYFVSHDFRDHAKNALLCRSVRTVKQMOV 373  
Search completed: February 27, 2002, 16:59:53  
Job time: 205 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 16:56:13 ; Search time 25.35 seconds  
(without alignments)  
1268.158 Million cell updates/sec

Title: US-08-816-011f-63  
Perfect score: 2294  
Sequence: 1 MVLINRNTYAVEQEAFPRD.....LTGGLWPVVIEHVDKPSIL 434

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2290	99.8	434	17	AA92315 CORK potassium cha
2	372.5	16.2	475	21	AA92286 Rat EN1 receptor
3	343.5	15.0	456	19	AAW69555 Human equilibrativ
4	343.5	15.0	456	21	AA155320 Human EN1 protein
5	327.5	14.3	475	21	AA92285 Human EN1 recepto
6	326.5	14.2	456	19	AAW83965 Equilibrative nucl
7	326.5	14.2	456	19	AAW69558 Human equilibrativ
8	326.5	14.2	456	21	AA15517 Human NEMPR-iENTP
9	326.5	14.2	475	21	AA939363 Human PRO1380 (UNQ
10	326.5	14.2	475	22	AAW66112 Protein of the inv
11	320.5	14.0	456	19	AAW69557 Rat equilibrative

12	301	13.1	457	19	AAW69556 Rat equilibrative
13	272.5	11.9	373	19	AAW64550 Human fibrosarcoma
14	261.5	11.4	397	22	AAW93550 Human protein sequ
15	214	9.3	326	21	AA15519 Human HNP36 protei
16	181	7.9	247	19	AAW83966 Alternatively spli
17	181	7.9	247	21	AA15518 Human NEMPR-iENTP
18	180	7.8	314	20	AAW88377 Polypeptide fragme
19	179.5	7.8	423	22	AAW88377 Human membrane or
20	119.5	5.2	383	22	AAW82466 S. epidermidis ope
21	114.5	5.0	527	21	AAW76126 Human secreted pro
22	113.5	4.9	329	22	AAW92687 Human protein sequ
23	113.5	4.9	360	22	AAW75546 Human secreted pro
24	113.5	4.9	530	22	AAW75598 Human secreted pro
25	113.5	4.9	545	22	AAW84667 Human protein sequ
26	113.5	4.9	579	21	AAW82460 Human SM-11044-bin
27	113.5	4.9	582	21	AAW82459 Human SM-11044-bin
28	111.5	4.9	397	21	AAW35641 Human PAR-2 protei
29	111.5	4.9	398	16	AAW66921 Human C140 recepto
30	111.5	4.9	398	17	AAW01953 Human C140 recepto
31	111.5	4.9	545	21	AAW94910 Human secreted pro
32	111	4.8	439	19	AAW61371 Non-adrenergic SM
33	108.5	4.7	776	21	AAW96167 Schizosaccharomyce
34	108	4.7	354	18	AAW25927 Xenopus melatonin
35	106.5	4.6	392	22	AAW06597 Human protein havi
36	106.5	4.6	399	16	AAW66922 Murine C140 recept
37	106.5	4.6	399	17	AAW01954 Murine C140 recept
38	106.5	4.6	439	22	AAW82657 S. epidermidis ope
39	106	4.6	318	22	AAW71666 Human olfactory re
40	106	4.6	318	22	AAW72405 Human OR-like poly
41	106	4.6	465	22	AAW95721 Human protein sequ
42	105.5	4.6	397	16	AAW66923 Human C140 recepto
43	105.5	4.6	397	17	AAW01955 Human C140 recepto
44	105.5	4.6	438	22	AAW60107 Human transport pr
45	105	4.6	393	21	AAW81642 Streptococcus pneu

ALIGNMENTS

RESULT 1  
AA92315 AAR92315 standard; Protein: 434 AA.  
XX  
AC AAR92315;  
XX  
DT 16-OCT-1996 (first entry)  
XX  
DE CORK potassium channel protein.  
XX  
KW CORK; potassium channel; nematode; pore-forming domain;  
KW transmembrane helix; N-glycosylation site; potassium-agonist;  
KW potassium-antagonist; drug screening; nematocide; anthelmintic;  
KW cardiac disorder.  
XX  
OS Caenorhabditis elegans.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 5..7 /note= "N-glycosylation site"  
FT Domain 26..39 /note= "Pore-forming H5 domain"  
FT Misc-difference 33..35 /note= "G-Y/F-G motif"  
FT Modified-site 81..83 /note= "N-glycosylation site"  
FT Modified-site 144..146 /note= "N-glycosylation site"  
FT Domain 150..162 /note= "Pore-forming H5 domain"  
FT Misc-difference 156..158 /note= "G-Y/F-G motif"  
XX  
PN W09613520-A1.

XX PD 09-MAY-1996.  
 XX PF 25-OCT-1995; 95WO-US14364.  
 XX PR 31-OCT-1994; 94US-0332312.  
 XX PA (AMCY ) AMERICAN CYANAMID CO.  
 XX PI Pausch MH, Price LA;  
 XX PW 1996-239450/24.  
 DR N-PSDB; AAT18168.  
 XX PT Potassium channel genes from Drosophila melanogaster and  
 PT Caenorhabditis elegans - useful in assaying substances to determine  
 PT effects on cell growth, and in inhibiting nematode and insect pests  
 XX  
 XX Claim 13; Fig 9; 79pp; English.  
 XX This potassium channel sequence is encoded by the CORK gene from  
 CC Caenorhabditis elegans, and has 2 pore-forming domains situated  
 CC between hydrophobic transmembrane helix domains. The CORK sequence  
 CC contains structural features resembling pore-forming H5 domains  
 CC found in potassium channels. 2 putative pore-forming H5 domains  
 CC contain the G-Y/F-G tripeptide motif required for potassium  
 CC selectivity. The protein contains 3 asparagine-linked glycosylation  
 CC sites. The protein may be expressed in a heterologous host cell to  
 CC assay substances to determine effects on cell growth. Potassium-  
 CC agonists or potassium-antagonists identified by this method may be  
 CC used as nematocides, anthelmintics or in therapy of cardiac  
 CC disorders, etc.  
 XX  
 XX Sequence 434 AA;

Query Match 99.8%; Score 2290; DB 17; Length 434;  
 Best Local Similarity 99.8%; Pred. No. 9.3e-228;  
 Matches 433; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVLINRNTAVAEQAFPRDKYNIVYLVILVGVLLPWNMTITTAPEYVYVWKPDPG 60  
 DB 1 MVLINRNTAVAEQAFPRDKYNIVYLVILVGVLLPWNMTITTAPEYVYVWKPDPG 60  
 QY 61 VETWYSKEFGSLTIGSLPNASINVFNLIIAGPLIYRVFAPVCFNIVNLIIILVI 120  
 DB 61 VETWYSKEFGSLTIGSLPNASINVFNLIIAGPLIYRVFAPVCFNIVNLIIILVI 120  
 QY 121 VLEPTDSMSWFFWTLGMAISINFSNGLYENSIVYGVGDFPHYTGALLIGNICGLLI 180  
 DB 121 VLEPTDSMSWFFWTLGMAISINFSNGLYENSIVYGVGDFPHYTGALLIGNICGLLI 180  
 QY 181 TVVKIGVYFLNDEPKLVAIVYGISLVILLVCAIALFFITKQDFYHHQKMEIREKA 240  
 DB 181 TVVKIGVYFLNDEPKLVAIVYGISLVILLVCAIALFFITKQDFYHHQKMEIREKA 240  
 QY 241 ETRDPSILMTWTFTCYGLFVNWVCFVATLTIFPVMVTVTTRGDSGLFNKIMSENDEI 300  
 DB 241 ETRDPSILMTWTFTCYGLFVNWVCFVATLTIFPVMVTVTTRGDSGLFNKIMSENDEI 300  
 QY 301 YTLTSLFVNLFAATIGSIVASIKHWPTRPYLKFAILRALFIPFFFCNRYVOTRAYPV 360  
 DB 301 YTLTSLFVNLFAATIGSIVASIKHWPTRPYLKFAILRALFIPFFFCNRYVOTRAYPV 360  
 QY 361 FFESTDIFVIGGTAMSFHGYLSALAMGYTPNVVPSHYSFAQLSVCTLMVGLLTGGLW 420  
 DB 361 FFESTDIFVIGGTAMSFHGYLSALAMGYTPNVVPSHYSFAQLSVCTLMVGLLTGGLW 420  
 QY 421 PVVIEHFVDKPSIL 434  
 DB 421 PVVIEHFVDKPSIL 434

RESULT 2  
 AAY82286  
 ID AAY82286 standard; Protein; 475 AA.  
 XX  
 AC AAY82286;  
 XX  
 DT 19-JUN-2000 (first entry)  
 XX  
 DE Rat EN1 receptor SEQ ID NO:5.  
 XX  
 KW Rat; EN1; EN1; immunosuppressant; vasotropic; thrombolytic;  
 KW cytotatic; hypotensive; antiinflammatory; analgesic; anticoagulant;  
 KW nucleoside transporter; ischaemia; cerebral embolism; malignant tumour;  
 KW organ transplant rejection; renitis; pancreatitis; hypertension;  
 KW analgesic; blood platelet coagulation inhibitor.  
 XX  
 OS Rattus sp.  
 XX  
 PN WO200012550-A1.  
 XX  
 PD 09-MAR-2000.  
 XX  
 PF 26-AUG-1999; 99WO-JP04602.  
 XX  
 PR 27-AUG-1998; 98JP-0241248.  
 XX  
 PA (KYOW ) KYOWA HAKKO KOGYO KK.  
 PI Miyaji H, Mimura H, Kambe M, Nakagawa S;  
 XX  
 DR WPI; 2000-256588/22.  
 DR N-PSDB; AAA08030.  
 XX  
 PT Novel nucleoside transporter polypeptide, useful in drug development  
 PT for agents in treating e.g. ischemia, cerebral embolism, rejection in  
 PT organ transplant, malignant tumors, renitis, pancreatitis and  
 PT hypertension  
 XX  
 PS Example 3; Page 81-83; 94pp; Japanese.  
 XX  
 CC The present invention describes human and rat EN1 receptors (EN1R1),  
 CC with some amino-acids deleted, substituted or added, but which retain  
 CC nucleoside transport activity, and can be used in drug development for  
 CC agents in treating e.g. ischaemia, cerebral embolism, rejection in organ  
 CC transplant, malignant tumors, renitis, pancreatitis and hypertension.  
 CC The EN1R1 proteins are nucleoside transporters which have vasotropic,  
 CC immunosuppressant, thrombolytic, cytotatic, hypotensive, analgesic,  
 CC antiinflammatory and anticoagulant activities. The transporter proteins,  
 CC their DNA, oligonucleotides, transformants and antibodies can be used  
 CC in drug development for agents in treating ischaemia, cerebral embolism,  
 CC rejection in organ transplant, malignant tumors, renitis, pancreatitis  
 CC and hypertension, and as analgesics, or blood platelet coagulation  
 CC inhibitors, as well as agents to reduce side-effects during  
 CC chemotherapy. The present sequence represents the rat EN1R1 protein.  
 XX  
 SQ Sequence 475 AA;

Query Match 16.2%; Score 372.5; DB 21; Length 475;  
 Best Local Similarity 25.6%; Pred. No. 7.5e-30;  
 Matches 120; Conservative 79; Mismatches 209; Indels 61; Gaps 13;

QY 7 SNTYAVEQAF-----PRDKYNIVYLVILVGVLLPWNMTITTAPEYV 52  
 DB 21 snrheadqallgkpldypagqlrpedrfngayiffclgllpwnffvt-akeywa 79  
 QY 53 NYWFK-----PDGVETWYSKEFGSLTIGSLPNASINVFNLIIAGPLIYRV 102  
 DB 80 ---fkircnsspasgkdpdadilnyfesyavastvpsllflvnlvnrivhrvl 136  
 QY 103 APVCFNIVNLTIILIVILEPTEDSMW---FFWVTLGMAISINFSNGLYENSIVYGVG 159  
 DB 137 aslsvalaifvnmavlr-----dtsswtrgrffsiamacmailssstfnssvyltg 191



therapy -  
 PS Disclosure; Fig 2A-B; 55pp; English.  
 CC The invention relates to the isolation of a gene encoding a NBMPR-iENTP  
 CC (nitrobenzylmercaptopyridine-2-yl)-insensitive, equilibrative nucleoside  
 CC transporter protein) from HeLa cells. The protein transports nucleosides  
 CC across the plasma membrane by a facilitated diffusion process and is  
 CC involved in the proliferative response. The transporter is useful in  
 CC screening to identify natural nucleoside permeants and/or their  
 CC inhibitors or analogues, potential therapeutic agents, also for studies  
 CC on protein structure and mechanism. Cells that express the iENTP as the  
 CC only transporter protein are used for drug screening (especially to  
 CC identify antitumor and antiviral nucleoside analogues), in chemotherapy  
 CC of cancer and for selective expression of heterologous genes for gene  
 CC therapy. The cDNA is used for recombinant expression of iENTP and as  
 CC a source of oligonucleotides (diagnostic primers and probes, ribozymes  
 CC and antisense sequences). Antibodies raised against iENTP are used for  
 CC detection of the protein by usual immunoassays and as (ant)agonists of  
 CC iENTP activity. This sequence represents the human ENT1 protein and  
 CC is used for comparison with the human NBMPR-iENTP protein (AAB15517).  
 CC  
 CC Sequence 456 AA;  
 PT Query Match 15.0%; Score 343.5; DB 21; Length 456;  
 PS Best Local Similarity 24.3%; Pred. No. 7e-27;  
 XX Matches 112; Conservative 77; Mismatches 212; Indels 59; Gaps 13;  
 CC  
 QY 18 PRDKNYVWLVILVGVGLLPMNMTTIAPEYVYVWFKPDGVE---TWYSKFMGSLT 74  
 Db 7 pqrkyavlllglglllpwnffmt-atqyftnrdmsqnvslvtaelksdaqasaa 65  
 QY 75 IGSOLP--NASINVFNLIIAGLLIYVFEA-----PVCENIVN-----LTITLI 117  
 Db 66 paalpernsisaifnnvmticamlplllfytlnsfhqrllpqsrvllgsvalllvfl 125  
 QY 118 LVIVLEPTDSMSFFFWVTLGMATSNFNSGLYNSVYGVGGDPFHTYIGALLIGNNICG 177  
 Db 126 tailkvqldalp-ffvitmkivlnsfagilqsgslfglagllpasytapimgsgqlag 184  
 QY 178 LLITVWIKGVTYFNDPKLVAIVYFGISLVILLVCAIALFFYKQDPYHYHQ----- 231  
 Db 185 ffasvami-calasgselsesafyfitacaviiltiicylgprlefyryyqqlieg 243  
 QY 232 -----KGMEIREKAEFDRPSILWTT-----FTNCYQLFNWVFCFAVT 271  
 Db 244 geqetkidliskgeepragkeesgvsvnsqptneshsikaiklnisvlfsvcfifit 303  
 QY 272 LTIPVVMVTVTRGDSGLFNKIMSENDEIYTLTSLVFNLFNFAAIGSIVASKIHWP--TP 329  
 Db 304 igmfpaevtvekvsslag-----sstweryfipvscfltnifdwlgrrsltavfmpgkds 358  
 QY 330 RYLFAILRALPIPFPPFCNRYVOTRAY-PVFPESTDIFVIGGIMSFSGHYSALAMG 388  
 Db 359 rwlpslvarlrvfplllcn--ikprytlvtvfehdawfifmaafafngylaslcmc 416  
 QY 389 YTPNVPVSHYSRAAQSVCTLMWGLLTGGGLWVPVIEHFV 428  
 Db 417 fgpkkvpaetaetagaiaffclgialgavsfilfraiv 456  
 RESULT 5  
 AAY82285  
 ID AAY82285 standard; Protein; 475 AA.  
 AC AAY82285;  
 XX  
 DT 19-JUN-2000 (first entry)  
 XX Human ENT1 receptor  
 DE Human ENT1 receptor SEQ ID NO:1.  
 XX Human; ENT1; ENT1; immunosuppressant; vasotropic; thrombolytic;

KW cytostatic; hypotensive; antiinflammatory; analgesic; anticoagulant;  
 KW nucleoside transporter; ischaemia; cerebral embolism; malignant tumour;  
 KW organ transplant rejection; renitis; pancreatitis; hypertension;  
 KW analgesic; blood platelet coagulation inhibitor.  
 XX Homo sapiens.  
 OS WO200012350-A1.  
 PN 09-MAR-2000.  
 XX 26-AUG-1999; 99WO-JP04602.  
 PF 27-AUG-1998; 98JP-0241248.  
 PR (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA Miyaaji H, Mimura H, Kambe M, Nakagawa S;  
 PI WPI; 2000-256588/22.  
 XX N-PSDB; AAA08027.  
 DR Novel nucleoside transporter polypeptide, useful in drug development  
 PT for agents in treating e.g. ischaemia, cerebral embolism, rejection in  
 PT organ transplant, malignant tumors, renitis, pancreatitis and  
 PT hypertension  
 XX Claim 1; Page 70-74; 94pp; Japanese.  
 PS  
 XX The present invention describes human and rat ENT1 receptors (ENTR1),  
 CC with some amino-acids deleted, substituted or added, but which retain  
 CC nucleoside transport activity, and can be used in drug development for  
 CC agents in treating e.g. ischaemia, cerebral embolism, rejection in organ  
 CC transplant, malignant tumors, renitis, pancreatitis and hypertension.  
 CC The ENTR1 proteins are nucleoside transporters which have vasotropic,  
 CC immunosuppressant, thrombolytic, cytostatic, hypotensive, analgesic,  
 CC antiinflammatory and anticoagulant activities. The transporter proteins,  
 CC their DNA, oligonucleotides, transformants and antibodies can be used  
 CC in drug development for agents in treating ischaemia, cerebral embolism,  
 CC rejection in organ transplant, malignant tumors, renitis, pancreatitis  
 CC and hypertension, and as analgesics, or blood platelet coagulation  
 CC inhibitors, as well as agents to reduce side-effects during  
 CC chemotherapy. The present sequence represents the human ENTR1 protein.  
 CC  
 CC Sequence 475 AA;  
 Query Match 14.3%; Score 327.5; DB 21; Length 475;  
 Best Local Similarity 24.4%; Pred. No. 3.3e-25;  
 Matches 116; Conservative 83; Mismatches 202; Indels 75; Gaps 15;  
 QY 7 SNTYAVEQEA-----PRDKYNIVWLVILVGVGLLPMNMTTIAPEYV 52  
 Db 21 ssslradealleklldrrppglrpedrfcgtlyiffslgslpwnffit-akeywm 79  
 QY 53 NYWFK-----PDGVETWYSKFMGSLTIGSOLPNASINVFNLIIAGLLIYR 100  
 Db 80 ---fklrnspsatgedpsd--ilnyfesyavastvpsmclvanflivrvavhir 134  
 QY 101 VFAPVCNIVNLTILILVIVLEP--TEDSMW---FPWVTLGMATSNFNSGLYNSVY 155  
 Db 135 vla-----sltvilaifmvtalvkvtftswtrgffavtvcmvilsgastvfsssi 187  
 QY 156 GVGSGPHTYIGALLIGNNICGLITVYVKGIVTYFLNDEPKLVAIVYFGISLVILLVCAI 215  
 Db 188 gmtsgfpmrnsqalisgmggtvsavasi-vdlaasdsrvnsalffltatfivlcmg 246  
 QY 216 ALFFITTKODEFYHY-----HHQKGEIREKAEFDRPSISi-----LWTTFTN 256  
 Db 247 lyllslrlyaryymrpvlaahvfsgge--elpqdsisapsvasrfdshpplpkk 304  
 QY 257 CYGOLFNVWFCFAVTLTIFPVVMTVTRGDSGLFNKIMSE--NDEIYTLTSLVFNLF 314



Db 305 taslgfctvtyffitsliypavct-----nieslnksgslwttkfflptlttlllynfad 359

QY 315 ATGSIVASKIHWTP--RYLKFAIILRALFIPFEFFCNVRVQTRAYPVFFESTDIFVIGG 372

Db 360 lcgrqitawiqpgpnskalpsfvilrtclpflvclnqprvhkvtvgdsvypalls 419

QY 373 IAMSFSHGYSALAMGYTPNVVPSHSYRFAAQLSVCTLMVGLLTGGLWPVVIHEFV 428

Db 420 silglisngylstlallypkivprelacaatgvmsfyvcglgtlsacstllivhii 475

RESULT 6

AAW83965

ID AAW83965 standard; Protein; 456 AA.

XX AC AAW83965;

XX DT 08-FEB-1999 (first entry)

XX DE Equilibrative nucleoside transport protein (iENTP).

XX KW Equilibrative nucleoside transport protein; iENTP; NAMPR; transport;

KW nitrobenzylmercaptapurine riboside; antiviral; antitumour; screening;

KW inhibitor; drug; knockout mice; cancer; chemotherapy; gene therapy;

KW adenosine deaminase; factor VIII.

XX OS Homo sapiens.

XX FH Key

FT Domain 13..27 Location/Qualifiers

FT /note= "transmembrane domain TM1"

FT Domain 71..93

FT /note= "transmembrane domain TM2"

FT Domain 100..117

FT /note= "transmembrane domain TM3"

FT Domain 125..144

FT /note= "transmembrane domain TM4"

FT Domain 162..185

FT /note= "transmembrane domain TM5"

FT Domain 194..214

FT /note= "transmembrane domain TM6"

FT Domain 293..312

FT /note= "transmembrane domain TM7"

FT Domain 326..341

FT /note= "transmembrane domain TM8"

FT Domain 361..379

FT /note= "transmembrane domain TM9"

FT Domain 393..413

FT /note= "transmembrane domain TM10"

FT Domain 432..452

FT /note= "transmembrane domain TM11"

XX WO9846749-A1.

XX PN 22-OCT-1998.

XX PD 10-APR-1998; 98WO-US07283.

XX PF 09-APR-1998; 98US-0058389.

XX PR 11-APR-1997; 97US-0838845.

XX ST JUDE CHILDREN'S RES HOSPITAL.

XX PA Belt JA, Crawford CR, Patel DH;

XX PI WPI; 1998-594576/50.

XX DR N-PSDB; AAW69698.

XX PT New equilibrative nucleoside transport protein insensitive to

XX PT nitrobenzylthio-purine riboside - useful for, e.g. identifying

XX PT specific therapeutic nucleoside analogues and in gene therapy to

XX PT protect transduced cells against ablative chemotherapy

PS Claim 13; Pages 71-72; 114pp; English.

XX This represents an equilibrative nucleoside transport protein (iENTP),

CC which is insensitive to nitrobenzylmercaptapurine riboside (NBMPR). Cells

CC transformed with a construct containing the iENTP nucleic acid can be

CC used to produce the protein recombinantly. iENTP is used to identify

CC specific ligands (particularly antiviral and antitumour nucleoside

CC analogues that are preferentially transported into cells) and to raise

CC antibodies. Cells in which iENTP provides all available transport

CC activity are used: (a) to identify permeants of iENTP and (b) to screen

CC specific inhibitors of iENTP (potential drugs). Fragments of the iENTP

CC nucleic acid are used, as probes, primers, antisense molecules, and

CC ribozymes for therapy or diagnosis, and knockout mice in which both

CC alleles encoding iENTP contain an inactivating defect are also useful for

CC drug screening. Cells that have been transduced with iENTP nucleic acid

CC ex vivo are used particularly for cancer chemotherapy. Vectors in which

CC the iENTP gene is linked to a heterologous gene (e.g. encoding adenosine

CC deaminase and factor VIII) can be used for gene therapy.

XX Sequence 456 AA;

QY 18 PRDKNIVYVYLVLGFGVLLPNMFITTAPEY-----VNYWFKPDGVE 62

Db 7 prdsyhlvgisffilgcllpwnffitaipfqlarlagagstariitstnhtgpedaf- 65

QY 63 TWYSKEFGSLTIGSOLPNASINVFNLFIAGLIYRVFAPVCFNVLTIILIVIL 122

Db 66 -----nfnmwvllsqpllllftllnsflyqcvpetvrlg---sllaillfaltaal 116

QY 123 EPTEDSMWFFWTLGMATSFNSNGLYENSVYGGDFPHYIGALLIGNNICGLLITV 182

Db 117 vkvdmspgpfisitmasvcfinsfsavlqgsifglgtmpstystiflsgqglagifaa 176

QY 183 VKI-----GVTYFLNDEPKLVAIVYF---GISLVILLVCAIALFFITKODFY---HYHHQ 231

Db 177 amllsmasgv-----daetsalgyfitpcvgilmsivcylslphikfarylankssqa 230

QY 232 KGMEIREKA-----ETDRPSPSILWTTFTNC 257

Db 231 qaqeletkaellqsdengipspqkvalltdlekepesepeqkpkpsvftvfkil 290

QY 258 YGQLFNWVFCFAVTLTIFFVMMVTVTRGDSGFLNKIMSENDEIYLLTSLFVNFPAAG 317

Db 291 wltalcilvltvtlsvfpaitamvtsts-----pgkwsqffnfpiccllfnmdwig 344

QY 318 SIVASKIHWTP--TPRYLKFAIILRALFIPFEFFCNVRVQTRAYPVFFESTDIFVIGGIAM 375

Db 345 rsitsyflwpdedsrllpllvclrlfvlplmclhvpqrsr-lpifpqdayfitmlif 403

QY 376 SFSHGYSALAMGYTPNVVPSHSYRFAAQLSVCTLMVGLLTG 417

Db 404 avsnlylvsitmcclaprvlpherevagalmftfialgiscg 445

RESULT 7

AAW69558

ID AAW69558 standard; Protein; 456 AA.

XX AC AAW69558;

XX DT 13-OCT-1998 (first entry)

XX DE Human equilibrative nucleoside transporter 2.

XX KW Human; equilibrative nucleoside transporter; hENT1; hENT2; rENT1;

KW rENT2; coronary; cerebrovascular anoxia; viral infection; cancer.

XX OS Homo sapiens.

XX W09829437-A2.  
 XX  
 PD 09-JUL-1998.  
 XX  
 PF 30-DEC-1997; 97WO-IB01657.  
 XX  
 PR 03-NOV-1997; 97US-0064004.  
 PR 30-DEC-1996; 96US-0034083.  
 XX  
 PA (UYAL-) UNIV ALBERTA.  
 PA (UYLE-) UNIV LEEDS.  
 XX  
 PI Baldwin SA, Cass CE, Young JD;  
 XX  
 DR WPI; 1998-388035/33.  
 DR N-PSDB; AAV40279.  
 XX  
 PT Newly isolated equilibrative nucleoside transporter protein(s) and  
 PT gene(s) - used to develop products for treating disorder(s)  
 PT associated with the transporter(s) and for use with nucleoside  
 PT drug(s)  
 XX  
 PS Claim 5; Fig 20; 97pp; English.  
 XX  
 CC The present sequence represents a substantially purified equilibrative  
 CC nucleoside transporter (ENT), human ENT2 (hENT2). ENTs can transport a  
 CC variety of purines and pyrimidines, including adenosine, uridine,  
 CC guanosine, inosine, formycin B, tubercidin, and thymidine. ENTs are  
 CC bidirectional, they transport a suitable permeant both into and out of  
 CC cells. ENTs can be used as a tool for the development of new nucleoside  
 CC drugs. Products from the present invention can be used for treating a  
 CC subject having a disorder associated with an ENT. They can also be used  
 CC with nucleoside drugs in the treatment of e.g. coronary or  
 CC cerebrovascular anoxia, viral infection or cancer. The products (e.g.  
 CC antibodies and oligonucleotides) hybridising to nucleic acid sequences  
 CC encoding ENTs) can also be used for detection and drug screening.  
 XX  
 SQ Sequence 456 AA;

Query Match 14.2%; Score 326.5; DB 19; Length 456;  
 Best Local Similarity 22.9%; Pred. No. 4e-25;  
 Matches 106; Conservative 71; Mismatches 200; Indels 85; Gaps 11;

QY 18 PRDKNYVWLVILVGVLLPWNMTITIAPEYV-----VNWFKPDGVE 62  
 DB 7 prdshyhgisffilgtltpwnffitaipfqrilagaggnstarihtsthtpedaf- 65  
 QY 63 TWSKEFMGSLTIGSQLPNASINVENFLIITAGPLIYRVFAPVCFNIVNLTIILVIVL 122  
 DB 66 -----nfnmwvllsqplllftllnsflyqcvpetvrlg----sllaillfaltaal 116  
 QY 123 EPTEDSMWFFWTLGMATSNFNSGLYENSIVGVGDFPHYTGALLIGNICGLITV 182  
 DB 117 vkvdmspgpffsiltmasvcfnstnsavlqslfgqlgmpststflisqgglagifaa 176  
 QY 183 VKI-----GVTVFLNDPKLVAIVVF---GISVLVLVCAIALPFIKQDPY---HVHHQ 231  
 DB 177 amllsmasgv-----daetsalgyfitpyvgilmsivcylsiphikfarylankssqa 230  
 QY 232 KGMIREKA-----ETDRSPSILMTTPTNC 257  
 DB 231 qaqelctkaellsgdengipsspqkvaltdldiekepesepepqpqkpsvtfvqki 290  
 QY 258 YGQLFNWFCFAVTLTFPPVMMVTTRGDSGLFNKINSENDEIYTLTSLVENLFAAIG 317  
 DB 291 wltalcvlvftvcisvpaltamvts-----Pqkwsqffnplccflifnimdwlg 344  
 QY 318 SIVASKIHP--TPRYLKFAILLRALRPIPFPPFFFCNRYQVTRAYPVFFESTDIFVIGGIAM 375  
 DB 345 rslcysflwpderrllpllvrcflifvpfmlchvqrsr-lpillfqdayfitfmlf 403

QY 376 SFSHCYLSALAMGYTPNVVPSHYSRFAAQSVCTLMVGLLTG 417  
 DB 404 avsgylvslmclaprqvlpheevagalmtfflaglglscg 445

RESULT 8  
 AAB15517  
 ID AAB15517 standard; Protein: 456 AA.  
 XX  
 AC AAB15517;  
 XX  
 DT 14-FEB-2001 (first entry)  
 XX  
 DE Human NBMPR-iENTP protein.  
 XX  
 KW Antiviral; antitumour; NBMPR-iENTP; nitrobenzylmercaptopyrineriboside;  
 KW insensitive, equilibrative nucleoside transporter protein; cancer;  
 KW plasma membrane; facilitated diffusion; gene therapy; primer; probe.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US6130065-A.  
 XX  
 PD 10-OCT-2000.  
 XX  
 XX 09-APR-1998; 98US-0058389.  
 PF  
 XX 11-APR-1997; 97US-0043659.  
 PR  
 XX (SUUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
 PA  
 XX Crawford CR, Belt JA;  
 PI  
 XX WPI; 2000-637839/61.  
 XX N-PSDB; AAA95722.  
 DR  
 DR New nucleic acid encoding nucleoside transporter, useful for drug  
 PT screening to identify antitumour and antiviral agents and for gene  
 PT therapy  
 PT  
 XX Claim 1; Fig 2A-B; 55pp; English.  
 XX  
 CC This sequence represents nitrobenzylmercaptopyrineriboside-insensitive,  
 CC equilibrative nucleoside transporter protein (NBMPR-iENTP) isolated from  
 CC Hela cells. The protein transports nucleosides across the plasma membrane  
 CC by a facilitated diffusion process and is involved in the proliferative  
 CC response. The transporter is useful in screening to identify natural  
 CC nucleoside permeants and/or their inhibitors or analogues, potential  
 CC therapeutic agents, also for studies on protein structure and mechanism.  
 CC Cells that express the iENTP as the only transporter protein are used  
 CC for drug screening (especially to identify antitumour and antiviral  
 CC nucleoside analogues), in chemotherapy of cancer and for selective  
 CC expression of heterologous genes for gene therapy. The cDNA is used  
 CC for recombinant expression of iENTP and as a source of oligonucleotides  
 CC (diagnostic primers and probes, ribozymes and antisense sequences).  
 CC Antibodies raised against iENTP are used for detection of the protein  
 CC by usual immunoassays and as (ant)agonists of iENTP activity.  
 XX  
 SQ Sequence 456 AA;

Query Match 14.2%; Score 326.5; DB 21; Length 456;  
 Best Local Similarity 22.9%; Pred. No. 4e-25;  
 Matches 106; Conservative 71; Mismatches 200; Indels 85; Gaps 11;

QY 18 PRDKNYVWLVILVGVLLPWNMTITIAPEYV-----VNWFKPDGVE 62  
 DB 7 prdshyhgisffilgtltpwnffitaipfqrilagaggnstarihtsthtpedaf- 65  
 QY 63 TWSKEFMGSLTIGSQLPNASINVENFLIITAGPLIYRVFAPVCFNIVNLTIILVIVL 122  
 DB 66 -----nfnmwvllsqplllftllnsflyqcvpetvrlg----sllaillfaltaal 116



PR 03-NOV-1998; 98US-0106905.  
 PR 03-NOV-1998; 98US-0106919.  
 PR 03-NOV-1998; 98US-0106932.  
 PR 03-NOV-1998; 98US-0106934.  
 PR 10-NOV-1998; 98US-0107783.  
 PR 17-NOV-1998; 98US-0108775.  
 PR 17-NOV-1998; 98US-0108779.  
 PR 17-NOV-1998; 98US-0108787.  
 PR 17-NOV-1998; 98US-0108788.  
 PR 17-NOV-1998; 98US-0108801.  
 PR 17-NOV-1998; 98US-0108802.  
 PR 17-NOV-1998; 98US-0108806.  
 PR 17-NOV-1998; 98US-0108807.  
 PR 17-NOV-1998; 98US-0108867.  
 PR 17-NOV-1998; 98US-0108925.  
 PR 18-NOV-1998; 98US-0108848.  
 PR 18-NOV-1998; 98US-0108849.  
 PR 18-NOV-1998; 98US-0108850.  
 PR 18-NOV-1998; 98US-0108851.  
 PR 18-NOV-1998; 98US-0108852.  
 PR 18-NOV-1998; 98US-0108858.  
 PR 18-NOV-1998; 98US-0108904.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;  
 XX WPI: 2000-237871/20.  
 DR N-PSDB; AAA37045.  
 DR  
 XX New mammalian DNA sequences encoding transmembrane, receptor or  
 PT secreted PRO polypeptides, useful for screening of potential peptide or  
 PT small molecule inhibitors of the relevant receptor/ligand interactions  
 XX  
 XX Claim 12; Fig 48; 773pp; English.  
 PS  
 CC AAA37022 to AAA37144 encode the new isolated human transmembrane,  
 CC receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The  
 CC transmembrane and receptor PRO proteins can be used for screening of  
 CC potential peptide or small molecule inhibitors of the relevant  
 CC receptor/ligand interactions. The polypeptides and nucleotide sequences  
 CC encoding then have various industrial applications, including uses as  
 CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent  
 CC PCR primers and hybridization probes used in the isolation of the PRO  
 CC polypeptides from the present invention.  
 XX  
 SQ Sequence 475 AA;  
 Query Match 14.2%; Score 326.5; DB 21; Length 475;  
 Best Local Similarity 24.4%; Pred. No. 4.2e-25;  
 Matches 116; Conservative 83; Mismatches 202; Indels 75; Gaps 15;  
 QY 7 SNTYAVEQAEF-----PRDKYNIYVWLVLGVFGVLLPWNFITIAPEYV 52  
 Db 21 ssslradealleklldrrppgqlrpedrcgtylffsiglsilpwnffit-akeywm 79  
 QY 53 NYWFK-----PDGVETWYSKEFMGSLTIGSQLPNASINVPNLFLLIAGLIYR 100  
 Db 80 ---fklrnsstpatgedpegsd--ilnyfesylavastvpsmiclvnfilnrvavhir 134  
 QY 101 VFAPVCNIVNLITILVILVLEP--TEDSMW---FFWVTLGMATYINFSNGLYNSVY 155  
 Db 135 vla-----slvtlaifnvtalvkvdtsswtrgffavtlvcmvilsgastvfsssy 187  
 QY 156 GVSGDPHTYIGALLIGNNICGLLITVVKYGVYFLNDEPKLVAVYVFGISLVLLVCAI 215  
 Db 188 gmtgspmrnsqaliggamggtvsavas1-vdlaessdvrsalafllatflvlcmg 246  
 QY 216 ALFFITKQDFYHY-----HHQKGMREIKAEATDRPSPSI-----LWTTFTN 256  
 Db 247 lyllslrlearymrvpvaahvfsgae--elpqdsisapsvasrfidshstplrpilkk 304

QY 257 CYGOLFNVWFCFAVTLTFIPVMMVTTRGDSGFLNKIMSE--NDEIYTLTSLFVFNLFA 314  
 Db 305 taslgfcvtyvffitsliypavct-----nieslnksgsgslwtckfflptlctfllynfad 359  
 QY 315 AIGSIVASKIHWPYP--RYLKFAILLRALFIPFFFFCNRYVQTRAYPVFFESTDIFVIGG 372  
 Db 360 lcgrqltawiqvpgpnskaipgfvllrtclilpflvclnyqprvhlktvfgsdvypalls 419  
 QY 373 IAMSFSHGYSALAMGYTPNVVPSHYSRFAQLSVCTLMVGLLTLGGLWPVVEHFV 428  
 Db 420 sliglsngyistlallygpixivprelaeatgvmsfyvcigltlgsacstllvhl 475  
 RESULT 10  
 AAB66112  
 ID AAB66112 standard; protein; 475 AA.  
 XX  
 AC AAB66112;  
 XX  
 DT 02-APR-2001 (first entry)  
 XX  
 DE Protein of the invention #24.  
 XX  
 KW Secreted; transmembrane; gene therapy.  
 XX  
 OS Unidentified.  
 XX  
 PN WC200078961-A1.  
 XX  
 PD 28-DEC-2000.  
 XX  
 PF 18-FEB-2000; 2000WO-US04342.  
 XX  
 PR 23-JUN-1999; 99US-0141037.  
 PR 20-JUL-1999; 99US-0144758.  
 PR 28-JUL-1999; 99US-0145698.  
 PR 01-SEP-1999; 99WO-US20111.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 05-JAN-2000; 2000WO-US00219.  
 PR 06-JAN-2000; 2000WO-US00376.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;  
 PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;  
 PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;  
 PI Watanabe CK, Williams PM, Wood WI;  
 XX  
 DR WPI: 2001-071395/08.  
 XX  
 PT Secreted and transmembrane proteins and nucleic acids designated PRO,  
 PT useful as hybridization probes, in chromosome and gene mapping and gene  
 PT therapy -  
 XX  
 PS Claim 1; Fig 48; 787pp; English.  
 XX  
 CC The present invention relates to secreted and transmembrane proteins.  
 CC These proteins and the DNA encoding them may be used as hybridization  
 CC probes, in chromosome and gene mapping and in the generation of  
 CC anti-sense RNA and DNA. They may also be used to generate either  
 CC transgenic animals or knockout animals which are in turn useful for  
 CC development and screening of therapeutically useful reagents.  
 CC The nucleic acids may also be used in gene therapy.  
 XX  
 SQ Sequence 475 AA;

Query Match 14.2%; Score 326.5; DB 22; Length 475;  
 Best Local Similarity 24.4%; Pred. No. 4.2e-25;  
 Matches 116; Conservative 83; Mismatches 202; Indels 75; Gaps 15;



```

XX 03-NOV-1997; 97US-0064004.
PR 30-DEC-1996; 96US-0034083.
XX
XX (UYAL-) UNIV ALBERTA.
PA (UYLE-) UNIV LEEDS.
XX
XX Baldwin SA, Cass CE, Young JD;
PI
XX
XX WPI: 1998-388035/33.
DR N-PSDB; AAV40277.
XX
XX Newly isolated equilibrative nucleoside transporter protein(s) and
PT gene(s) - used to develop products for treating disorder(s)
PT associated with the transporter(s) and for use with nucleoside
PT drug(s)
XX
XX Claim 5; Fig 9; 97pp; English.
XX
XX The present sequence represents a substantially purified equilibrative
CC nucleoside transporter (ENT), rat ENT1 (rENT1). ENTs can transport a
CC variety of purines and pyrimidines, including adenosine, uridine,
CC guanosine, inosine, formycin B, tubercidin, and thymidine. ENTs are
CC bidirectional, they transport a suitable permeant both into and out of
CC cells. ENTs can be used as a tool for the development of new nucleoside
CC drugs. Products from the present invention can be used for treating a
CC subject having a disorder associated with an ENT. They can also be used
CC with nucleoside drugs in the treatment of e.g. coronary or
CC cerebrovascular anoxia, viral infection or cancer. The products (e.g.
CC antibodies and oligonucleotides hybridising to nucleic acid sequences
CC encoding ENTs) can also be used for detection and drug screening.
XX
XX Sequence 457 AA;
SQ

```

Query Match 13.1%; Score 301; DB 19; Length 457;  
 Best Local Similarity 22.3%; Pred. No. 1.7e-22;  
 Matches 107; Conservative 84; Mismatches 191; Indels 98; Gaps 17;

```

QY 18 PRKYNIVWLIVGVLLPNWMTTIAPEYVYVWFKPDGV-----ETWYSKFFMGS 72
Db 7 pdcyaxwllfvlgltlpwnffit-atqyftsrlnsqisvntqscstaealad 65
QY 73 LTGSGOLPNASINVENFLIIAGLIYRVFAPVCFN-----IVNLTII 115
Db 66 psvslparssalsafnnvmtlcamlplif--tclnsflhqkvsgslrllgslallvlf 123
QY 116 LILVIVLEPTEDSMWFFWTLGCMATSNFNSGLYNSVYGVGDDPHTYVIGALLIGNNI 175
Db 124 lvtativkvqmdals-fftiitmikivlinsfgallqasifglagvipanytapimgsgql 182
QY 176 CGLLITVVKI-----GVTYFLNDEPKLVAIVYFGISLVILLVCAIALFFFTKQ 223
Db 183 agftsvamicavsgskisesafyfit---acavv-----ilaicyalpwm--- 229
QY 224 DFYHHQ-----KGMEIR-EKATDRPSPIL----- 250
Db 230 efyrhylqlnlagpaegketkldiisegeprgreesvgpgpnlpanrnqskailksl 289
QY 251 WTTFTNCGOLFNVWFCFVTLTIFPVMVTVTTRGDSGLFNKMTSENDEITYTLTSLVLF 310
Db 290 wvl-----alsvcfiftvtlglfp---avtaevessiaagtspwknc-yfipvacflnf 338
QY 311 NLEFAAGISIVASKIHP--TPRYLKFAILRALFIFPFFFCNRYVOTRAYVPVFESTDIF 368
Db 339 nvfdwlgrrsltaicmwpqgdsrwlplvacrvvfiplmlcnvk-qhhylpslfdkdwvf 397
QY 369 VIGGIAMSFSGHLSALAMGYTPNVVPSHSYRFAALSVCTLMVGLLTGGLWPVVIEHFV 428
Db 398 itmaefatsngylaslcmcfcpkpkvpaetaetagnimsfflclglalgavlsllralv 457

```

RESULT 13

```

AAW64550
ID AAW64550 standard; Protein; 373 AA.
XX
AC AAW64550;
XX
DT 21-OCT-1998 (first entry)
XX
XX Human fibrosarcoma cell line HT-1080 clone HP10235 protein.
XX
XX Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW differentiation; immune system; stimulator; suppressor; regulator;
KW haematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour.
XX
OS Homo sapiens.
XX
XX WO9821328-A2.
XX
XX 22-MAY-1998.
XX
XX 07-NOV-1997; 97WO-JP04056.
XX
XX 13-NOV-1996; 96JP-0301429.
XX
XX (PROT-) PROTEGENE INC.
XX (SAGA ) SAGAMI CHEM RES CENTRE.
XX
XX Kato S, Kobayashi M, Sekine S, Yamaguchi T;
XX WPI: 1998-297932/26.
XX N-PSDB; AAV49582, AAV49583.
XX
XX Human protein having transmembrane domain - useful for, e.g.
PT research and nutrition
XX
XX Claim 1; Page 108-109; 205pp; English.
XX
XX AAW64534-W64558 represent human proteins containing a transmembrane
CC domain. These proteins can be used for, e.g. research and nutrition, and
CC may have cytokine and cell proliferation/differentiation, immune
CC stimulating/suppressing, haematopoiesis regulating, tissue growth,
CC activin/inhibin, chemotactic/chemokinetic, haemostatic and thrombolytic,
CC receptor/ligand, anti-inflammatory or tumour inhibition activity.
XX
XX Sequence 373 AA;
SQ

```

Query Match 11.9%; Score 272.5; DB 19; Length 373;  
 Best Local Similarity 23.6%; Pred. No. 1.1e-19;  
 Matches 91; Conservative 66; Mismatches 187; Indels 41; Gaps 9;

```

QY 73 LTGSGOLPNASINVENFLIIAGLIYRVFAPVCFNIVNLTIIILVIVLEPTEDSMWF 132
Db 1 mticamlplllfynslhqrpqsvrlgsl---vallvflitalivkvqldalp-f 56
QY 133 FWTGLGATSNFNSGLYNSVYGVGDDPHTYVIGALLIGNNICGLLITVVKIGVYFLN 192
Db 57 fvitmikivlinsfgallqsglflaglipasytapimgsgglagffasvami-calasg 115
QY 193 DEPKLVAIVYFGISLVILLVCAIALFFFTKQDEYHHQ-----KGME 235
Db 116 selseafyfitacavviltiicyglprletfryyqqllegpgeqetkldliskge 175
QY 236 IREKATDRPSPSILWTT-----FTNCGOLFNVWFCFVTLTIFPVMVTVTTRGD 286
Db 176 pragkeesgvsvnsqptsheshikailnislafsvcfiftitigmfpavtvekssi 235
QY 287 SGLFNKIMSENDEITYTLTSLVFLNFAAGISIVASKIHP--TPRYLKFAILRALFIP 344
Db 236 ag-----sstweryfipvscfltnifdwlgrrsltavfmwpgkdswlpslvarlvfp 290
QY 345 FFFFCNRYVOTRAY-DVFFESTDIFVIGGIAMSFSGHLSALAMGYTPNVVPSHSYRFAA 403

```

Db 291 llllcn--ikpryltvvfehdawffmaafangylaslcmcfgpkkvkpaetaag 348  
Qy 404 OLSVCTLMVGLLTGGLWPVIEHFV 428  
Db 349 aimaffclglgavfifraiv 373  
RESULT 14  
AAB93550  
ID AAB93550 standard; Protein; 397 AA.  
XX AAB93550;  
AC AAB93550;  
DT 26-JUN-2001 (first entry)  
DE Human protein sequence SEQ ID NO:12928.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
OS Homo sapiens.  
PN EP1074617-A2.  
XX EP1074617-A2.  
PD 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 8; SEQ ID 12928; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesising 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 397 AA;

Query Match 11.4%; Score 261.5; DB 22; Length 397;  
Best Local Similarity 23.7%; Pred. No. 1.7e-18;  
Matches 92; Conservative 72; Mismatches 181; Indels 43; Gaps 10;  
Qy 69 FMGSLTIGSOLPNASINVFNLFLIITAGPLIYRVFAPVCNIVNLTILILVIVLEP--TE 126  
Db 25 fesyIavasavpsmclclvanflvravhrlva-----sltvlaifmvtalvkv 77  
Qy 127 DSMW---FFVTLGMAISNFSGLYSVGVGDPDPTTYIGALLIGNNICGLLITVV 183  
Db 78 dtsswtrgfavtvcmwilsgastvfssiygmctgspmrnsqalisggamgtvsava 137  
Qy 184 KIGVTYFLNDEPKLVAIVYFGISLVILVCAIALFFITKQDFYHY-----HHOKGME 235  
Db 138 sl-vdlaasdvrrnsalaffltatflvcmglylllsrleyarymrvpvlaahvfsgge 196  
Qy 236 IREKAETORPSPSI-----LWTFPTNCYQOLFNNWFCFAVTLTIFPVMVTVTR 284  
Db 197 --elpqdsisapsvasrfdidshpplrpilkkktaslgfcvtyvffitsliypavct---- 250  
Qy 285 GDSGFLNKIMSE--NDEIYTLTSLVFNLFRAIGSIVASKIHWPDP--RYLKFAILRA 340  
Db 251 -niesinksgsgslwttkfipttfllynfadlcrqltawrvpnpnskalpgfvilrt 309  
Qy 341 LFIPEFFFCNRYVQTRAYPVFPFESTDIFVIGGIAMSFSGYLSALAMGYTPNVVPSHSYR 400  
Db 310 clipflvcnypvrhltktvfgsdvypallssllsglsglystiallygpkivprelae 369  
Qy 401 FAAQLSVCTLMVGLLTGGLWPVIEHFV 428  
Db 370 atgvvmstfyvclgtlgsaacstllvhlh 397  
RESULT 15  
AAB15519  
ID AAB15519 standard; Protein; 326 AA.  
XX AAB15519;  
AC AAB15519;  
DT 14-FEB-2001 (first entry)  
XX Human HNP36 protein.  
DE Human HNP36 protein.  
XX  
KW Antiviral; antitumour; NBMPR-1ENTP; nitrobenzylmercaptopyrineriboside;  
KW insensitive, equilibrative nucleoside transporter protein; cancer;  
KW plasma membrane; facilitated diffusion; gene therapy; primer; probe.  
XX  
OS Homo sapiens.  
XX US6130065-A.  
PN 10-OCT-2000.  
PD 09-APR-1998; 98US-0058389.  
PF 11-APR-1997; 97US-0043659.  
PR (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
XX Crawford CR, Belt JA;  
PI WPI; 2000-637839/61.  
XX New nucleic acid encoding nucleoside transporter, useful for drug  
PT screening to identify antitumour and antiviral agents and for gene  
PT therapy -  
XX Disclosure; Fig 2A-B; 55pp; English.  
XX The invention relates to the isolation of a gene encoding a NBMPR-1ENTP  
CC (nitrobenzylmercaptopyrineriboside-insensitive, equilibrative nucleoside  
CC transporter protein) from HeLa cells. The protein transports nucleosides  
CC across the plasma membrane by a facilitated diffusion process and is

CC involved in the proliferative response. The transporter is useful in  
 CC screening to identify natural nucleoside permeants and/or their  
 CC inhibitors or analogues, potential therapeutic agents, also for studies  
 CC on protein structure and mechanism. Cells that express the iENTP as the  
 CC only transporter protein are used for drug screening (especially to  
 CC identify antitumour and antiviral nucleoside analogues), in chemotherapy  
 CC of cancer and for selective expression of heterologous genes for gene  
 CC therapy. The cDNA is used for recombinant expression of iENTP and as  
 CC a source of oligonucleotides (diagnostic primers and probes, ribozymes  
 CC and antisense sequences). Antibodies raised against iENTP are used for  
 CC detection of the protein by usual immunoassays and as (ant)agonists of  
 CC iENTP activity. This sequence represents the human HNP36 protein and  
 CC is used for comparison with the human NBMPR-iENTP protein (AAB15517).  
 XX  
 SQ Sequence 326 AA;

Query Match 9.3%; Score 214; DB 21; Length 326;  
 Best Local Similarity 22.0%; Pred. No. 1e-13;  
 Matches 71; Conservative 52; Mismatches 139; Indels 60; Gaps 8;  
 QY 143 INFSGLYENSVGVGDFPHYIGALLIGNNIGLLITVVKI-----GVITYFLNDEPKL 197  
 Db 7 insfsavlgqsfqglqcmptystlflsgqglagifaafamllsmasgv-----daet 60  
 QY 198 VAIYVF---GISLVILLVCAIALEFFITKQDFY---HYHHOKGMEIREKA----- 240  
 Db 61 saigyfitpyvgilmsivcylslphlkfarylankssqqaqaeletkaellqsdengip 120  
 QY 241 -----ETDRSPSILMTTFTNCTNYGQLFNWVFCFAVTLTTFPV 277  
 Db 121 sspqkvaltidldlekesepdepqpgkpsvftvfqkiwltalcilvlvftvltsvfa 180  
 QY 278 MMTVTTRGDSGLNKNINSENDEIYTLTSLFVNLFRAAGSIVASKIHWP--TPRYLKEA 335  
 Db 181 itamvtsts-----pgkwsqffnplccflfnimdwigrsltsyflwpdedsrllp 234  
 QY 336 IILRALFPFFFCNRYVQTRAVPVFFESTDIFVIGIAMSFSGHLSALAMGYTPNVVP 395  
 Db 235 vclrlflvpflmchvpgsr-lpilfpqdayfitmllfavsngylvslmtclaprql 293  
 QY 396 SHYSRFAAQLSVCTLMVGLLTG 417  
 Db 294 pherevagalmftflalglscg 315

Search completed: February 27, 2002, 16:59:32  
 Job time: 199 sec